

Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie und Pharmazie
der Ludwig-Maximilians-Universität München

**Incorporation of pyrrolysine derivatives into proteins
and development of bioorthogonal protein
modification methods**

Michael Johannes Gattner

aus München

2018

Erklärung

Diese Dissertation wurde im Sinne von §7 der Promotionsordnung vom 28. November 2011 von Herrn Prof. Dr. Thomas Carell betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

Mannheim, den 21. Juli 2018

Michael J. Gattner

Dissertation eingereicht am 27. Dezember 2018

1. Gutachter: Prof. Dr. Thomas Carell
2. Gutachterin: Prof. Dr. Anja Hoffmann-Röder

Mündliche Prüfung am 21. Februar 2019

Parts of this thesis were published:

Publications:

Gattner, M. J., Vrabel, M. & Carell, T. Synthesis of ϵ -N-propionyl-, ϵ -N-butyryl-, and ϵ -N-crotonyl-lysine containing histone H3 using the pyrrolysine system. *Chem. Commun.* **49**, 379–381 (2013).

Schneider, S., Gattner, M. J., Vrabel, M., Flügel, V., López-Carrillo, V., Prill, S. & Carell, T. Structural insights into incorporation of norbornene amino acids for click modification of proteins. *ChemBioChem* **14**, 2114–2118 (2013).

Vrabel, M., Kölle, P., Brunner, K. M., Gattner, M. J., López-Carrillo, V., de Vivie-Riedle, R. & Carell, T. Norbornenes in inverse electron-demand Diels-Alder reactions. *Chemistry* **19**, 13309–13312 (2013).

Gattner, M. J., Ehrlich, M. & Vrabel, M. Sulfonyl azide-mediated norbornene aziridination for orthogonal peptide and protein labeling. *Chem. Commun.* **50**, 12568–12571 (2014).

Ehrlich, M., Gattner, M. J., Viverge, B., Bretzler, J., Eisen, D., Stadlmeier, M., Vrabel, M. & Carell, T. Orchestrating the biosynthesis of an unnatural pyrrolysine amino Acid for its direct incorporation into proteins inside living cells. *Chemistry* **21**, 7701–7704 (2015).

Datz, S., Argyo, C., Gattner, M., Weiss, V., Brunner, K., Bretzler, J., Schirnding, von, C., Torrano, A. A., Spada, F., Vrabel, M., Engelke, H., Bräuchle, C., Carell, T. & Bein, T. Genetically designed biomolecular capping system for mesoporous silica nanoparticles enables receptor-mediated cell uptake and controlled drug release. *Nanoscale* **8**, 8101–8110 (2016).

Further publication:

Flühe, L., Knappe, T. A., Gattner, M. J., Schäfer, A., Burghaus, O., Linne, U. & Marahiel, M. A. The radical SAM enzyme AlbA catalyzes thioether bond formation in subtilisin A. *Nat. Chem. Biol.* **8**, 350–357 (2012).

Danksagung

An dieser Stelle möchte ich einigen Menschen danken, die zum Gelingen dieser Arbeit beigetragen haben.

Zuallererst danke ich meinem Doktorvater Prof. Thomas Carell, der mir die Möglichkeit gab, in seiner Arbeitsgruppe in einem spannenden Forschungsgebiet zu arbeiten. Ich bedanke mich für die vielseitige Themenstellung, interessante Diskussionen und Anregungen, das entgegengebrachte Vertrauen und die Bereitstellung exzellenter Arbeitsbedingungen. Die hier erlernten Fähigkeiten haben es mir erlaubt, meinen angestrebten Weg als Wissenschaftler in der Industrie zu gehen.

Ich danke den Mitgliedern meiner Prüfungskommission für die Evaluierung meiner Arbeit. Bei Frau Prof. Dr. Hoffmann-Röder möchte ich mich herzlich für die Übernahme der Zweitkorrektur bedanken.

Für die Unterstützung bei organisatorischen Belangen bedanke ich mich bei Slava Gärtner, Sabine Voß und Kerstin Kurz.

Für das sorgfältige Korrekturlesen dieser Arbeit bedanke ich mich ganz herzlich bei Dr. Veronica Lopez Carrillo, Dr. Markus Müller, Dr. Michael Ehrlich und Dr. Milan Vrabel.

Ein großer Dank geht an Dr. Markus Müller für seine Hilfe und Unterstützung in allen biochemischen Fragestellungen. Außerdem bedanke ich mich bei ihm, bei Dr. Emine Kaya und bei Dr. Christian Deiml für die Einführung in das spannende Thema.

Besonders bedanken möchte ich mich bei Dr. Milan Vrabel, der mich stets mit frisch synthetisierten unnatürlichen Aminosäuren versorgt hat. Durch wissenschaftliche Diskussionen und sein unglaubliches Durchhaltevermögen inspirierte und motivierte er mich und trug maßgeblich zu dieser Arbeit bei. Milan, ich wünsche Dir weiterhin alles gute auf dem steinigen Weg zum Professor.

Auch bei Dr. Michael Ehrlich möchte ich mich herzlich für die Zusammenarbeit bedanken. Die Synthese des D-Ornithin-Derivats hatte sich als herausfordernder

herausgestellt, als wir ursprünglich annahmen. Es freut mich, dass wir das Projekt gemeinsam zum erfolgreichen Ende brachten.

Ich danke Dr. Christian Argyo und Dr. Stefan Datz und deren Gruppenleiter Prof. Thomas Bein für die gute Zusammenarbeit.

Allen Mitgliedern der Arbeitsgruppe danke ich für eine ausgesprochen angenehme Atmosphäre. Es hat Spaß gemacht, Teil dieser Gruppe zu sein. Nicht nur die gute Zusammenarbeit im Labor, sondern auch viele außeruniversitäre Unternehmungen bleiben mir im Gedächtnis. Da waren feuchtfröhliche Bootsausflüge, ausgiebige Wiesenbesuche, üppige Weißwurstfrühstücke, ausgelassene Caferamparties und gemütliche Feierabendbiere, um nur einige zu nennen. Auch das wöchentliche Fußballspiel zusammen mit einigen Kollegen aus dem Genzentrum war eine willkommene Ablenkung vom meistens aufregenden, manchmal aber auch frustrierenden Laboralltag. Besonders bedanken möchte ich mich bei meinen ehemaligen Arbeitskreis-Kollegen, jetzt Freunden Dr. Dorothea Matschkal, Dr. Ines Pfaffeneder, Dr. Sandra Koch, Dr. Felix Gnerlich, Dr. Martin Münzel, Dr. Michael Ehrlich, Dr. Toni Pfaffeneder, Dr. Caterina Brandmeyer, Dr. Stefan Prill, Dr. Mirko Wagner, Dr. Michael Ehrlich und Dr. Milan Vrabel.

Ein besonderer Dank gebührt meinen Eltern und meiner Schwester, die mich im Studium und in meiner Promotionszeit aber auch danach moralisch und finanziell unterstützt haben und mich auch danach immer wieder motiviert haben, die Arbeit abzuschließen.

Der letzte und größte Dank geht an Dr. Veronica Lopez Carrillo, meine ehemalige Kollegin, jetzt Verlobte und Mutter unserer wunderbaren Tochter Inara. Danke, dass Du Dich für einen PostDoc-Aufenthalt in der Carell-Gruppe entschlossen hast und Danke für Deine Unterstützung während des Anfertigens dieser Arbeit. Te quiero bonita!

Table of Contents

1	Summary	13
2	Zusammenfassung.....	16
3	Abbreviations	19
4	Introduction.....	22
4.1	Natural expansions of the genetic code.....	22
4.1.1	The 20 canonical amino acids.....	22
4.1.2	Selenocysteine, the 21st proteinogenic amino acid.....	23
4.1.3	Pyrrolysine, the 22nd proteinogenic amino acid	26
4.2	Incorporation of NCAAs into proteins	31
4.2.1	Applications of unnatural amino acids in proteins	31
4.2.2	Site-specific, genetically encoded incorporation of NCAAs.....	35
4.3	Bioorthogonal modifications of unnatural pyrrolysine homologs.....	46
4.3.1	Bioorthogonality and practical aspects of protein labeling strategies	46
4.3.2	Ketone-hydrazide/hydroxylamine reactions.....	47
4.3.3	Staudinger ligations	48
4.3.4	Azide-alkyne reactions.....	50
4.3.5	Inverse electron-demand Diels-Alder cycloadditions.....	52
4.3.6	Desmethylpyrrolysine-aminobenzaldehyde ligations	54
4.3.7	Other ligation methods	55
5	Literature of the introduction.....	56
6	Aims of the thesis	74

7	Original publications	75
7.1	Synthesis of ϵ -N-propionyl-, ϵ -N-butyryl-, and ϵ -N-crotonyl-lysine containing histone H ₃ using the pyrrolysine system	76
7.2	Norbornenes in inverse electron-demand Diels-Alder reactions.....	80
7.3	Structural insights into incorporation of norbornene amino acids for click modification of proteins.....	85
7.4	Sulfonyl azide-mediated norbornene aziridination for orthogonal peptide and protein labeling.....	91
7.5	Orchestrating the biosynthesis of an unnatural pyrrolysine amino Acid for its direct incorporation into proteins inside living cells	96
7.6	Genetically designed biomolecular capping system for mesoporous silica nanoparticles enables receptor-mediated cell uptake and controlled drug release.....	101
8	Contributions	112
9	Selected supplementary information of the original publications	114
9.1	Selected supplementary information corresponding to section 7.1.....	115
9.2	Selected supplementary information corresponding to section 7.4.....	137
9.3	Selected supplementary information corresponding to section 7.5	147

I Summary

Site specific incorporation of non-canonical amino acids (NCAAs) into proteins is a powerful tool for e.g. studying the effects of post translational modifications, revealing enzymatic mechanisms, visualizing proteins in living cells or adding new functionalities to therapeutic and diagnostic proteins. The technology is based on orthogonal tRNA/aminoacyl-tRNA synthetase (aaRS) pairs, which do not interfere with the biosynthetic machinery of the host cell (e.g. *E. coli*).

The tRNA/aaRS pair, which was used in this thesis originates from the archaeal strain *Methanosarcina mazei*. It is responsible for the ribosomal incorporation of pyrrolysine (Pyl), also known as the 22nd proteinogenic amino acid, in response to UAG ("amber-") codons. In the classic genetic code, these amber codons are stop codons and lead to translation termination. When transferring the tRNA^{Pyl}/PylRS-pair to e.g. *E. coli*, the incorporation of a NCAA in response to an amber codon competes with translation termination and is therefore called "amber suppression".

In this thesis, the pyrrolysine system was exploited to incorporate three newly described posttranslational lysine modifications site specifically into histone H3, where they have been identified. The amino acids ϵ -N-propionyl-, ϵ -N-butyryl-, and ϵ -N-crotonyl-lysine were directly incorporated at the desired position to simulate the respective naturally occurring propionylation, butyrylation or crotonylation respectively. Comprehensive bioanalytical examinations of the modified histone, which included peptide mapping-MS and western blot analyses unambiguously proved the site specific and complete incorporation of these NCAAs.

Although numerous NCAAs have already been incorporated into proteins by amber suppression, the structural diversity of these is of course limited. However, the introduction of NCAAs carrying bioorthogonal modification handles opens an infinite field of possibilities since it allows post translational, specific modifications of the protein of interest. One of the most promising handles are norbornenes, which could be introduced site specifically into proteins by the Carell group.¹ The modification of

norbornenes with tetrazines in inverse electron demand Diels-Alder reactions is one of the fastest bioorthogonal modification reactions described.

Mechanistic studies in collaboration with the de Vivie-Riedle group revealed and explained that, due to a more stable transition state, *exo*-substituted hydroxymethyl norbornenes react around three-times faster than their *endo* substituted isomers.

To analyze, whether this behavior could also be observed on the protein level, the *endo* substituted norbornyl lysine and two novel *exo* substituted norbornene amino acids were incorporated into model proteins. The acceptance of these new norbornene amino acids by the Pyl system including a PylRS triple mutant was supported with solid analytical data. The correct incorporation and complete labeling with fluorescent and PEGylated tetrazines was demonstrated by SDS-PAGE, intact-MS and peptide mapping analyses of the unmodified and modified protein species. A faster labeling reaction of the *exo*-substituted norbornyl amino acids was not observed in these *in vitro* protein labeling studies.

In collaboration with the Schneider group, crystal structures of the used PylRS triple mutant bearing the Pyl analogs demonstrated the versatility of the triple mutant developed in the Carell group.¹

In order to expand the set of bioorthogonal modification reactions, a novel click reaction, the aziridation of norbornenes using electron-deficient sulfonyl azides was developed. The reaction was characterized on small molecule level, on the peptide level and finally on the protein level using norbornene containing model proteins. The second order rate constant was determined to be $k_2 = 1.7 \times 10^{-3} \text{ M}^{-1} \text{ s}^{-1}$. This value is comparable to the rate constants of the strain-promoted azide-alkyne cycloaddition of the first generation which is commonly used for biomolecule labeling applications.^{2,3} Chemoselective protein conjugation with fluorescent, and immuno tags on the norbornene residue was proven by SDS-PAGE, Western blot as well as by peptide mapping and intact mass spectrometric studies. The reaction proceeds efficiently under mild conditions, does not require any catalysis and is orthogonal to functional groups of native proteins. Furthermore, sulfonyl azides are easily accessible compounds, which makes the newly developed labeling reaction an attractive alternative to the existing set of bioconjugation techniques.

A general challenge of the incorporation of NCAs into proteins is that the amino acids have to be chemically synthesized in large quantities to reach the required millimolar concentrations during protein expression. Our idea was to circumvent these laborious syntheses and hand over the work to the biosynthetic machinery of the host cell. In this thesis, a Pyl-biosynthesis "hijacking" system was developed, that allowed the biosynthesis and incorporation of unnatural 3*S*-ethynylpyrrolysine (ePyl), a bifunctional reaction handle, by feeding *E. coli* cells with 3*R*-ethynyl-D-ornithine. The correct production and incorporation of the novel NCAA in a purified model protein was proved by extended analytics including mass spectrometric studies on intact and peptide level. It was shown, that site specifically incorporated ePyl can be modified by two different orthogonal click reactions at only one amino acid site providing access to highly modified proteins.

Finally, to prove the feasibility of NCAA carrying proteins as useful tools in biomedical applications, site specifically modified *human carbonic anhydrase*-folic acid-conjugates (HCA-FA) were applied as targeting, pH responsive capping proteins for drug delivery in mesoporous silica-based nanocarriers (MSNs). The studies were carried out in collaboration with the groups of Prof. Bein and Prof. Bräuchle. Three key elements of the newly developed capping system arise from the site-specifically modified HCA-FA. First, the protein is large enough to block the gates of the MSN to prevent the cargo from escaping. Second, HCA binds phenyl-sulfonamide (phSA) and therefore sulfonamide-functionalized MSNs (MSN-phSA) in a pH depended manner. At pH values below 7, typically present during endocytosis of MSNs, the enzyme releases MSN-phSA, which ensures the release of the cargo at the exact right moment. Third, the site-specific functionalization of HCA can direct the MSNs to a specific cell type.

2 Zusammenfassung

Der ortsspezifische Einbau von nicht-kanonischen Aminosäuren (engl. abgekürzt NCAs) in Proteine ist ein leistungsfähiges Werkzeug in der chemischen Biologie. Es kann unter anderem für die Untersuchung der Auswirkungen von posttranslationalen Modifikationen, für die Aufklärung enzymatischer Mechanismen, für die Visualisierung von Proteinen in lebenden Zellen oder zum Funktionalisieren therapeutischer und diagnostischer Proteine verwendet werden. Die Technologie basiert auf orthogonalen tRNA/Aminoacyl-tRNA-Synthetase (aaRS)-Paaren, die nicht in die Biosynthesemaschinerie der Wirtszelle (z.B. *E. coli*) eingreifen. Das in dieser Arbeit verwendete tRNA/aaRS-Paar stammt aus dem Archaea-Stamm *Methanosarcina mazei*. Es ist verantwortlich für den ribosomalen Einbau von Pyrrolysin (Pyl), das auch als 22. proteinogene Aminosäure bekannt ist. Pyl wird vom sogenannten "Amber Codon" codiert. Im klassischen genetischen Code ist dieses Codon ein Stopp-Signal und führt zur Termination der Translation. Wenn das tRNA^{Pyl}/PylRS-Paar z.B. in *E. coli* exprimiert wird, konkurriert der Einbau einer NCA als Antwort auf ein Amber-Codon mit der Translationsterminierung und wird daher als *amber suppression* bezeichnet.

In dieser Arbeit wurde das Pyrrolysin-System genutzt, um drei neu beschriebene posttranslationale Lysin-Modifikationen spezifisch an Positionen in Histon H₃ zu integrieren, an denen sie identifiziert wurden. Die Aminosäuren ϵ -N-Propionyl-, ϵ -N-Butyryl- und ϵ -N-Crotonyllysin wurden direkt an der gewünschten Position eingebaut, um die jeweilige natürlich vorkommende Propionylierung, Butyrylierung bzw. Crotonylierung zu simulieren. Umfassende bioanalytische Untersuchungen des modifizierten Histons mit peptide mapping-MS und Western-Blot-Analysen belegten den ortsspezifischen und vollständigen Einbau dieser NCAs eindeutig.

Obwohl zahlreiche NCAs bereits durch *amber suppression* in Proteine eingebaut wurden, ist die strukturelle Vielfalt dieser Aminosäuren natürlich begrenzt. Die Einführung von NCAs, die bioorthogonal modifizierbare funktionale Gruppen tragen, eröffnet jedoch nahezu unerschöpfliche Möglichkeiten, da sie spezifische Modifikationen von Proteinen nach deren Expression ermöglicht. Eine der Vielversprechendsten dieser funktionalen Einheiten sind Norbornene, die von der Carell-Gruppe spezifisch in Proteine eingebaut

werden können.¹ Die Reaktion von Norbornenen mit Tetrazinen in invers Elektronen-anfordernden Diels-Alder-Cycloadditionen ist eine der schnellsten beschriebenen bioorthogonalen Modifizierungsreaktionen.

In mechanistischen Studien in Zusammenarbeit mit der Gruppe von Prof. de Vivie-Riedle konnte gezeigt und erklärt werden, dass *exo*-substituierte Hydroxymethyl-Norbornene aufgrund eines stabileren Übergangszustands etwa dreimal schneller mit Tetrazinen reagieren als ihre *endosubstituierten* Isomere. Um zu analysieren, ob dieses Verhalten auch auf Proteinebene beobachtet werden kann, wurden das *endo*-substituierte Norbornyllysine und zwei neue *exo*-substituierte Norbornen-Aminosäuren in Modellproteine eingebaut. Die Akzeptanz dieser neuen Norbornen-Aminosäuren durch das Pyl-System, das eine zuvor entwickelte PylRS-Dreifachmutante enthielt, wurde mit soliden analytischen Daten bewiesen. Der korrekte Einbau und die vollständige Markierung mit fluoreszierenden und PEGylierten Tetrazinen wurde durch SDS-PAGE-, Intakt-MS- und *peptide mapping*-Analysen der unmodifizierten und modifizierten Proteinspezies gezeigt. Eine schnellere Markierungsreaktion der *exo*-substituierten Norbornyl-Aminosäuren wurde bei diesen *in vitro*-Proteinmarkierungsstudien allerdings nicht beobachtet. In Zusammenarbeit mit der Arbeitsgruppe von Dr. Sabine Schneider konnte die Vielseitigkeit der in der Carell-Gruppe entwickelten PylRS Dreifachmutante in Kristallstrukturen mit gebundenen Pyl-Analoga gezeigt werden.¹

Um die Auswahl an bioorthogonalen Modifikationsreaktionen zu erweitern, wurde die Aziridierung von Norbornenen mit elektronenarmen Sulfonylaziden, eine neuartige Klickreaktion, entwickelt. Die Reaktion wurde auf niedermolekularer, auf Peptid- und schließlich auf Proteinebene mit Norbornenen, Norbornen-Peptiden und Norbornen-haltigen Modellproteinen charakterisiert. Die Geschwindigkeitskonstante zweiter Ordnung wurde mit $k_2 = 1,7 \times 10^{-3} \text{ M}^{-1} \text{ s}^{-1}$ bestimmt. Dieser Wert ist vergleichbar mit den Geschwindigkeitskonstanten der Ringspannungs-unterstützten Azid-Alkin-Cycloaddition der ersten Generation, die häufig für Biomolekül-Markierungen verwendet wird.^{2,3} Die chemoselektive Proteinkonjugation mit Fluoreszenz- und Immuno-reagenzien am Norbornen-Rest wurde durch SDS-PAGE, Western Blot sowie durch *peptide mapping* und intakte massenspektrometrische Untersuchungen nachgewiesen. Die Reaktion verläuft effizient unter milden Bedingungen, erfordert keine Katalyse und ist orthogonal zu

funktionellen Gruppen von nativen Proteinen. Darüber hinaus sind Sulfonylazide leicht zugängliche Verbindungen, was die neu entwickelte Markierungsreaktion zu einer attraktiven Alternative zu den bestehenden Biokonjugationstechniken macht.

Eine allgemeine Herausforderung beim Einbau von NCAs in Proteine besteht darin, dass die Aminosäuren in großen Mengen chemisch synthetisiert werden müssen, um die erforderlichen millimolaren Konzentrationen während der Proteinexpression zu erreichen. Unsere Idee war es, diese aufwendigen Synthesen zu umgehen und die Arbeit an die biosynthetische Maschinerie der Wirtszelle zu übergeben. Dafür wurde in diesem Projekt ein Pyl-Biosynthese-"Entführungs"-System entwickelt, das die Biosynthese und den Einbau von unnatürlichem β S-Ethynylpyrrolysin (ePyl), einer zweifach modifizierbaren Einheit, ermöglicht, indem *E. coli*-Zellen mit β R-Ethynyl-D-Ornithin gefüttert werden. Die korrekte Synthese und die richtige Position des Einbaus der neuen NCAA in einem gereinigten Modellprotein wurde durch umfassende Analysen einschließlich massenspektrometrischer Untersuchungen auf intakter und Peptidebene nachgewiesen. Es konnte gezeigt werden, dass ortsspezifisch eingebautes ePyl durch zwei unterschiedliche orthogonale Klickreaktionen modifiziert werden kann, was den Zugang zu hochmodifizierten Proteinen ermöglicht.

Um schließlich die Anwendbarkeit von NCAA-tragenden Proteinen als nützliche Werkzeuge in biomedizinischen Anwendungen zu beweisen, wurden ortsspezifisch modifizierte humane Carboanhydrase-Folsäure-Konjugate (HCA-FA) als zielgerichtete, pH-schaltbare *Capping*-Proteine für die Wirkstoffabgabe in mesoporösem Siliciumdioxid Nanoträgern (engl. abgekürzt MSNs) eingesetzt. Die Studien wurden in Zusammenarbeit mit den Gruppen von Prof. Bein und Prof. Bräuchle durchgeführt. Drei Schlüsselemente des neu entwickelten Verschlusssystems ergeben sich aus dem ortsspezifisch modifizierten HCA-FA. Erstens ist das Protein groß genug, um die Poren der MSNs zu blockieren, um zu verhindern, dass die Ladung austritt. Zweitens bindet HCA Phenylsulfonamid (phSA) und daher Sulfonamid-funktionalisierte MSNs (MSN-phSA) pH-abhängig. Bei pH-Werten unter 7, die typischerweise während der Endozytose von MSNs vorliegen, lässt das Enzym MSN-phSA los und gewährleistet damit die Freisetzung der Fracht zum genau richtigen Zeitpunkt. Drittens steuert die ortsspezifische Funktionalisierung von HCA die Anbindung der MSNs an spezifische Zelltypen.

3 Abbreviations

aaRS	Aminoacyl-tRNA synthetase
ADC	Antibody-drug-conjugate
AmpR	Ampicilin resistance
CHO	Chinese hamster ovary cells
CuAAC	Cu(I)-catalyzed azide-alkyne 1,3-dipolar cycloaddition
D-Orn	D-ornithine
DTT	Dithiotreithol
H ₃	Histone 3
HRP	Horse raddish peroxidase
iEDDA	Inverse electron demand Diels-Alder cycloaddition
INN	International nonproprietary name
IPTG	Isopropyl- β -thiogalactopyraniside
MALDI	Matrix assisted laser desorption/ionization
<i>Mj</i> TyrRS/ <i>Mj</i> tRNA ^{Tyr}	<i>Methanococcus jannaschii</i> tRNA/tRNA-aminoacyl synthetase pair
MMA	Monomethylamine
mRNA	messenger ribonucleic acid
MS	mass spectrometry

MtbB	Bimethylamine methyltransferase
MtmB	Monomethylamine methyltransferase
MttB	Trimethylamine methyltransferase
NCAA	Non canonical amino acid
Norb	Nobornene
Orn	Ornithine
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PLP	Pyridoxal phosphate
PTM	Post translational modification
Pyl	Pyrrolysine
PylRS	Pyrrolysyl-tRNA synthetase
PylT	Pyrrolysine-tRNA
rDA	Reverse Diels-Alder reaction
RF	Release factor
RP-HPLC	Reverse phase high performance liquid chromatography
SDS	Sodium dodecylsulfate
Sec	Selenocystein
SECIS	Selenocysteine insertion element

SerRS	Serine-tRNA synthetase
tRNA ^{aa}	tRNA specific for amino acid aa
tRNA ^{Pyl}	Pyrrolysine-tRNA
TrxA	Thioredoxin A
UAA	Unnatural amino acid
wt	Wild type

4 Introduction

4.1 Natural expansions of the genetic code

The set of 20 canonical amino acids has been seen as an unchangeable part of a universal genetic code. In contrast to this dogma, observed deviations of the genetic code and especially the discovery of a 21st and a 22nd proteinogenic amino acid have proven the mutability of the genetic code and have led to discussions about its origin.⁴

4.1.1 The 20 canonical amino acids

All organisms living on this planet share the same amino acid alphabet, even though their evolutionary pathways may have diverged very early in the development of life. The bacterium *Escherichia coli*, a sunflower and the human species, for instance, share the same genetic code, which decodes the same set of protein building blocks. But why did exactly these amino acids develop and why exactly 20?

A widely accepted hypothesis assumes that the first organisms were based entirely on ribonucleotides.⁵ The usage of only four building blocks enabled these ribo-organisms to form so called ribozymes. These are complex secondary structures which are able to catalyze specific chemical reactions. The following extension of the number of enzyme building blocks from four (ribo-organisms) to 20 (modern organisms) certainly increased the diversity of chemical functionality and therefore enabled more complex structures.

Nevertheless, enlargement of versatility cannot be the only driving force in the evolutionary development of amino acids.^{6,7} Several canonical amino acids show redundant functional properties as the acidic aspartate and glutamate, the amide containing asparagine and glutamine or the non-polar valine, leucine and isoleucine. A possible argument for the reduced complexity of the selected amino acid library could be their metabolic price.^{8,9} Bulky, more complex amino acids, such as phenylalanine or tyrosine are found in lower amounts than easily accessible amino acids as valine or serine. Based on this economical principle, only small and metabolically cheap amino acids found their way into the set of protein building blocks by the cost of lower versatility. Although, when only arguing economically, the metabolic costs could be even more reduced, since it was shown, that most of the known protein folding structures are accessible with much smaller sets of

amino acids.^{10,11} Therefore, the question why exactly those 20 amino acids were selected early in the development of life cannot entirely be explained as a compromise between gained complexity and low metabolic costs.

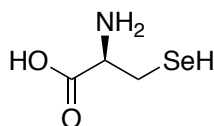
An explanation to this question is given by the so-called "frozen accident theory" first described by Francis Crick in 1968.¹² This theory describes a situation in which many organisms with different sets of amino acids existed. One of these organisms developed an innovation (accident) that favored it in a way that no other organism could compete anymore. Every change of the amino acid set would have made this innovation impossible to realize. Therefore, the set of amino acids was frozen and not a target of evolution anymore.

In contrast to this theory, natural genetic code variations such as stop-codon-read-through or the two additional amino acids selenocysteine (Sec) and pyrrolysine (Pyl) were found in natural proteins.¹³⁻¹⁶ Especially the discoveries of the additional amino acids Sec and later Pyl were stimulating the discussions about the origin of the canonical amino acids again.¹⁷

4.1.2 Selenocysteine, the 21st proteinogenic amino acid

Selenocysteine (Sec, U) was discovered in 1976 as the first proteinogenic amino acid which was not part of the standard set of 20 amino acids and is therefore often referred to as the "21st" proteinogenic amino acid.¹³ It was identified in all three domains of life but is most abundant in bacteria, where a quarter of all sequenced strains is synthesizing Sec.^{7,8} Even in the human genome, numerous Sec containing proteins, so called selenoproteins, were identified.¹⁸

The structure of Sec can be formally described as a cysteine analog of which the thiol group is exchanged by a selenol group (Scheme 4.1).



Scheme 4.1 Structure of selenocysteine (Sec)

Sec was shown to be specifically introduced in response to the opal stop codon (UGA), first described for *E. coli*.¹⁹⁻²¹ As depicted in Figure 4.1, several factors have to exist in a bacterial genome to be able to encode Sec. First, a special tRNA^{Sec} has to be present which recognizes the UGA codon. This tRNA^{Sec} is charged by a standard SerRS with serine to yield Ser-tRNA^{Sec} and is afterwards converted to Sec-tRNA^{Sec} by Sela, the selenocysteine synthase.^{22,23} Sela is a pyridoxal phosphate dependent enzyme which transfers selenium from selenium phosphate (provided by SelD) to tRNA^{Sec}-coupled serine. The synthesis of Sec exclusively occurs on tRNA^{Sec}-coupled serine, not on the free amino acid, which fundamentally differentiates the Sec incorporation pathway from the later discussed pyrrolysine incorporation pathway. SelB, a specific GTP dependent elongation factor binds the completed Sec-tRNA^{Sec} and a crucial mRNA stem loop structure in a defined distance downstream of the gene of interest. This mRNA secondary structure element is called SECIS: Selenocysteine insertion sequence.²⁴⁻²⁷ The complex of SelB, SECIS and Sec-tRNA^{Sec} enables the ribosome to translate an opal codon (UGA) by binding the complementary anti-codon loop of Sec-tRNA^{Sec} and inserting selenocysteine into the nascent protein chain.

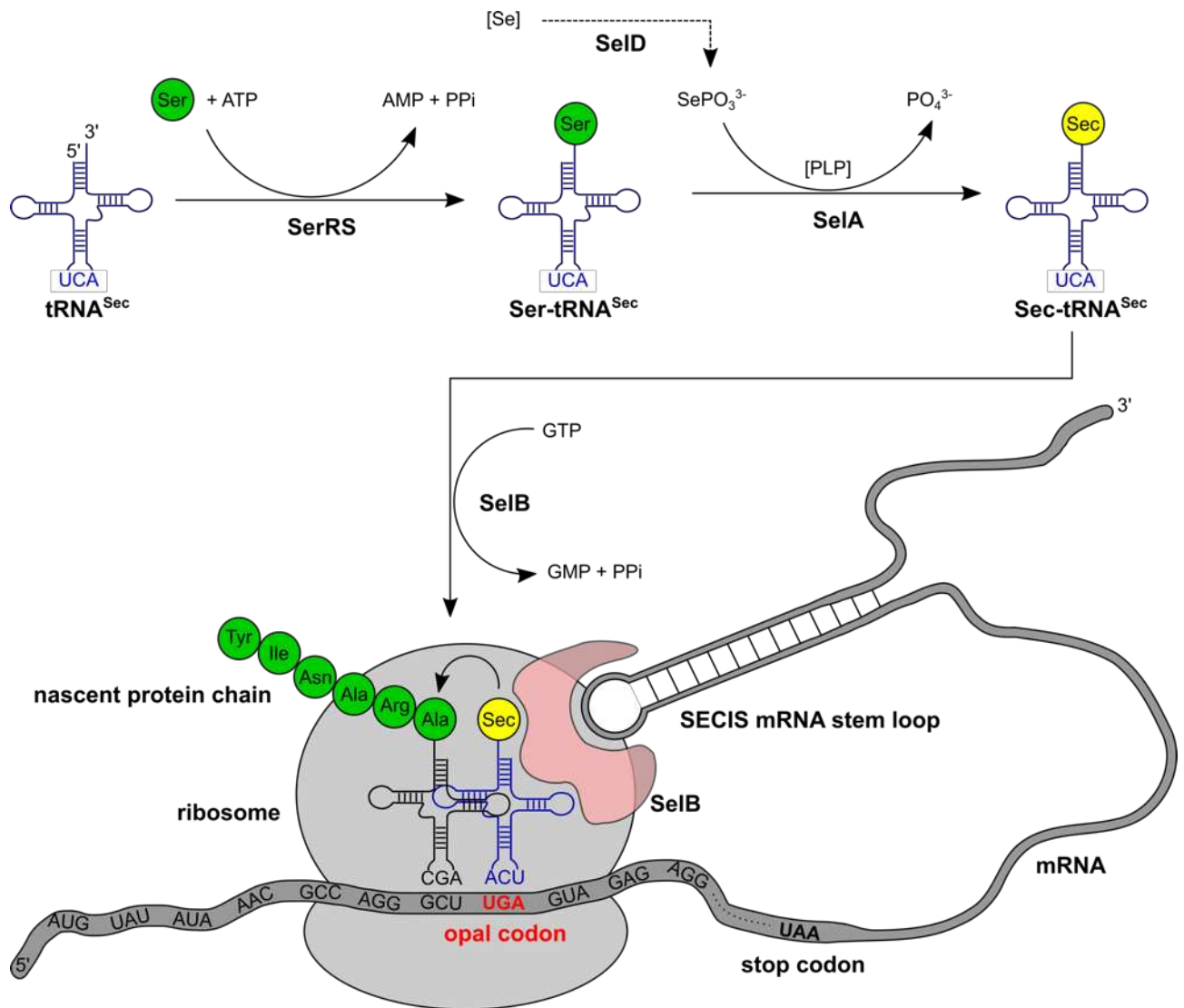


Figure 4.1 Selenocysteine synthesis and incorporation into proteins in prokaryotic cells

Analogues of SelB, SelD and SECIS have been identified in archaea and eukaryotes, where the principle incorporation of Sec works similarly, whereas the exact mechanisms of Sec biosynthesis differ slightly from the described bacterial pathway.^{28,29}

4.1.3 Pyrrolysine, the 22nd proteinogenic amino acid

The existence of pyrrolysine (Pyl, O) was first described in two independent reports in 2002 by Chan et al.¹⁴ and Krzycki et al.¹⁵ **Figure 4.2** shows the structure of pyrrolysine. Pyl contains a 4-methylated-(4R, 5R)-pyrroline-5-carboxylate, which is linked via an amide bond to the ϵ -nitrogen of L-lysine.³⁰⁻³²

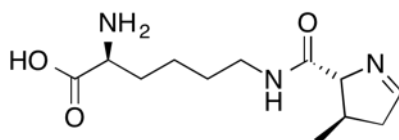


Figure 4.2 Structure of pyrrolysine (Pyl)

4.1.3.1 Natural function and distribution

The genetic ability to biosynthesize Pyl was discovered in more than 25 anaerobic archaea and bacteria. pyrrolysine containing proteins have been identified mainly in methanogenic archaea species of the family *Methanosarcinaceae* such as *Methanosarcina mazei* or *Methanosarcina barkeri*, but also in a few bacteria as in the bacteria *Desulfitobacterium hafniense* or *Bilophila wadsworthia* and in a symbiotic deltaproteobacterium of the gutless worm *Olavius algarvensis*.^{15,33,34}

Pyl enables the methanogenic *Methanosarcina* species to metabolize monomethyl-, dimethyl- and trimethylamine since it is expressed in the active sites of the methyltransferases MtmB, MtbB and MttB.^{31,35} Although these three enzymes show little structural homologies, the function of Pyl in their active sites is similar.³⁶⁻³⁸ Figure 4.3 exemplifies its function for the methyl transfer of monomethylamine by the methyltransferase MtmB: The first step of the methyl transfer is the addition of the methylamine to protonated Pyl, which activates the methyl group for transfer to MtmC, a cobalt-containing corrinoid cofactor bound enzyme. At the end of this enzymatic cascade, this methyl group is converted either to methane or to carbon dioxide or it enters pathways leading to carbon assimilation. Therefore, Pyl acts like a cofactor, with the difference, that it is incorporated cotranslationally as an amino acid.

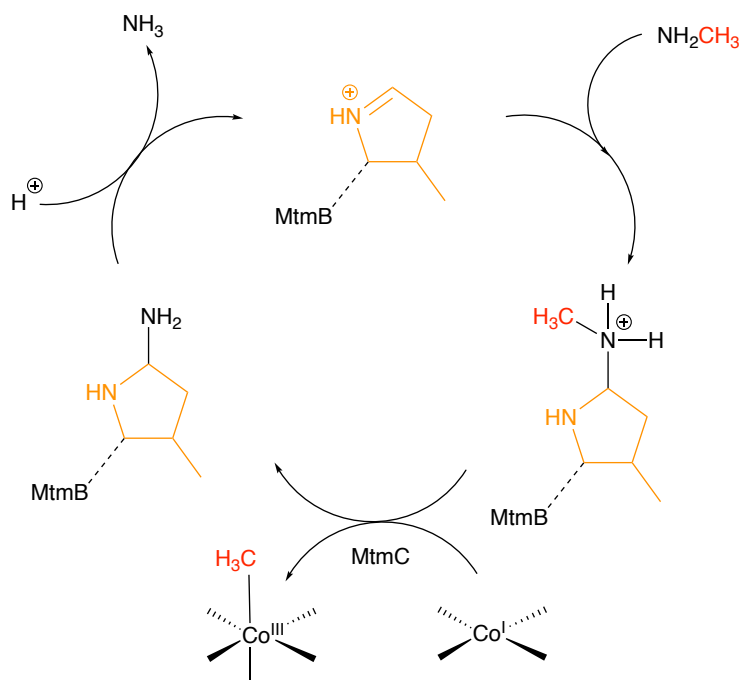


Figure 4.3 Transfer of a methyl group (red) from monomethylamine to Pyl (orange) in the catalytic site of MtmB

The function of bacterial Pyl containing methyltransferase homologues is not solved yet and their presence in the bacterial proteome probably is a result of horizontal gene transfer.³⁹

4.1.3.2 Biosynthesis

The Pyl gene cluster contains five genes: *pylT*, *pylS*, *pylB*, *pylC* and *pylD*. A transfer of only these five genes together with the *mtmB* gene enabled *E. coli* to generate functional MtmB containing Pyl in its active site.⁴⁰ These genes can therefore be considered as a genetic code expansion cassette.^{40,41} The genes *pylT* and *pylS* encode for an amber codon specific tRNA and its corresponding aminoacyl transferase PylRS which will be described in detail in the next chapter. The gene products of *pylB*, *pylC* and *pylD* are responsible for the biosynthesis of Pyl. The first attempts to identifying the underlying mechanisms suggested D-ornithine (D-Orn), **Figure 4.4**, as a possible precursor of Pyl, since it was shown, that the addition of this non-proteinogenic amino acid increased the yields of MtmB in the described heterologous *E. coli* expression system dramatically.⁴²

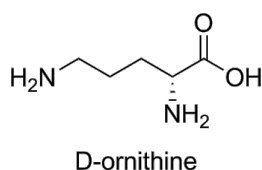


Figure 4.4 *Structure of D-ornithine (D-Orn)*

The exact mechanisms of the biosynthesis were revealed independently by the groups of Krzycki and Geierstanger in the year 2011, nine years after the discovery of Pyl.^{15,43,44} Stable isotope-labeling experiments showed, that the only precursor of Pyl formation is the amino acid L-lysine.⁴³ D-Ornithine was shown to take a shortcut in the biosynthetic pathway, which leads to the demethylated Pyl analog desmethylpyrrolysine (dmPyl), also called pyrroline-carboxy lysine (Pcl).⁴⁴ Figure 4.5 shows the three steps from two lysine molecules to Pyl.

The enzymes act in their genetical order from PylB to PylD. The formation of Pyl starts with the rearrangement of L-lysine to (3R)-3-methyl-D-ornithine. This impressive step is fulfilled in a radical reaction by PylB, a member of the iron-sulfur cluster containing radical SAM superfamily.⁴⁵ The subsequent amide bond formation between the carboxyl group of methyl ornithine and the ϵ -amino group of a second L-lysine is catalyzed by PylC, an ATP dependent ligase with structural similarities to ATP grasp superfamily enzymes.⁴⁶ The final step of the biosynthesis of Pyl is carried out by PylD, a dehydrogenase which harbors NAD⁺ as the coenzyme in its active site.⁴⁷ The enzyme catalyzes an oxidation of the δ -amino group of the fused ornithine moiety and a subsequent pyrroline ring formation.

The elucidation of the Pyl biosynthesis also explained the observed higher MtmB titers of D-ornithine fed, Pyl-gene cluster and *mtmB* carrying *E. coli*, which was described above:^{42,44} The first reaction step of the described biosynthesis (catalyzed by PylB) is the kinetic bottleneck of Pyl formation. This step can be circumvented by adding D-ornithine to the growth medium. D-ornithine is a homologue of the natural intermediate (3R)-3-methyl-D-ornithine. The only difference is a missing methyl group in position 3. Both of the following biosynthetic enzymes (PylC and PylD) as well as the amino acyl synthetase PylRS accept the resulting intermediates with a missing methyl group.

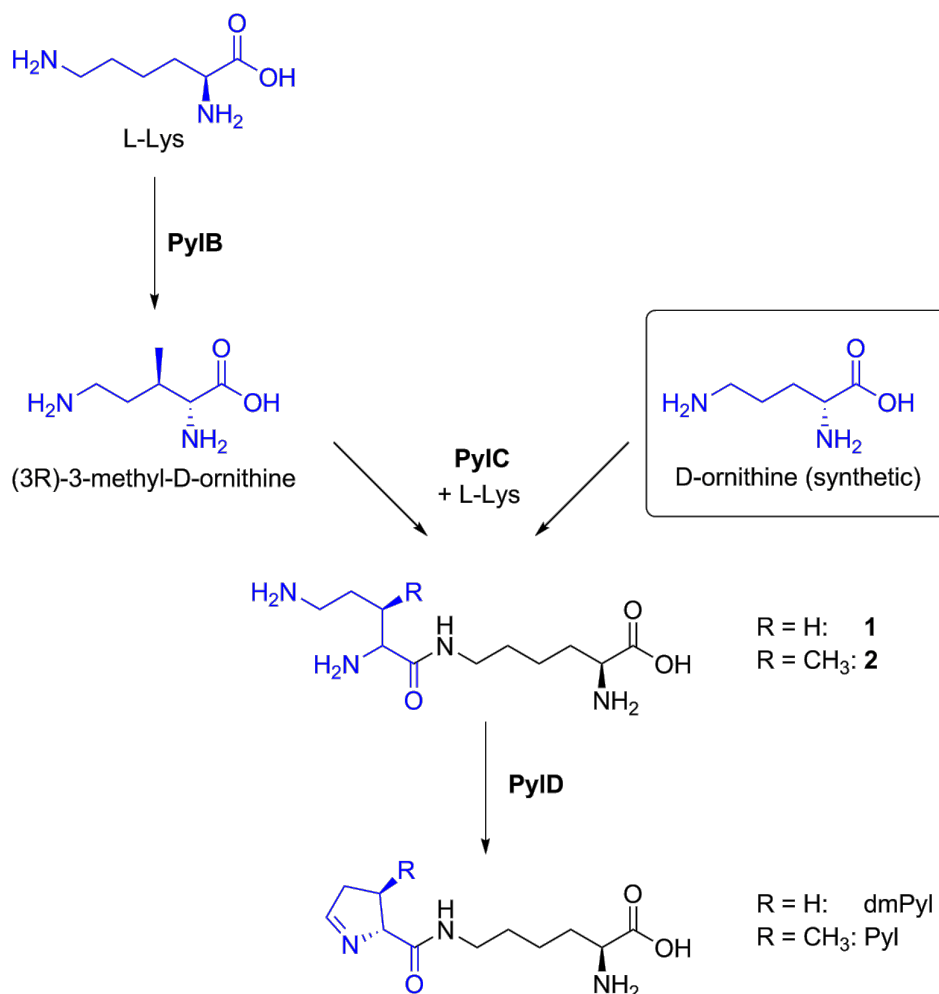


Figure 4.5 Biosynthesis of pyrrolysine (Pyl) from two L-lysines and artificial shortcut of D-ornithine to desmethylpyrrolysine (dmPyl)

4.1.3.3 Translation of the UAG amber codon to pyrrolysine

Adding Pyl to the amino acid set requires surprisingly few remarking adaptations of the conserved genetic code.

Since every triplet-mRNA codon is already occupied by either an amino acid or a stop signal during translation, reprogramming of one codon is needed. In case of Pyl, the affected codon is the mRNA triplet UAG, the so-called amber stop codon. In Pyl-containing organisms, this codon is no longer recognized as a stop signal but as a specific codon for Pyl. Therefore, in these organisms the signal to stop protein synthesis is coded by one of the two remaining stop codons, namely UAA (ochre codon) or UGA (opal codon).

The actual recoding of UAG by Pyl-expressing organisms is enabled by only three components: The amino acid Pyl itself, an amber specific tRNA carrying the triplet 5'-CUA-3' in its anticodon loop (tRNA^{Pyl}) and the corresponding amino acyl-tRNA synthetase (PylRS) which charges the tRNA^{Pyl} with Pyl.¹⁵ As shown in Figure 4.6, the charged tRNA^{Pyl} can enter the A site of the ribosome during protein synthesis and introduce Pyl into the nascent protein chain in response to an amber codon.

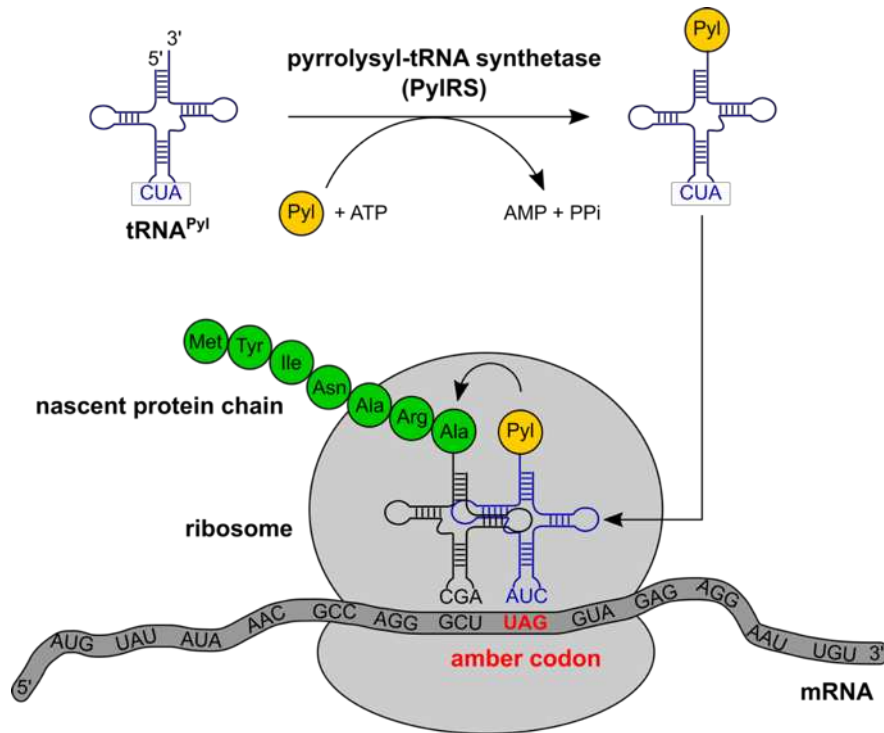


Figure 4.6 Charging of tRNA^{Pyl} by PylRS and incorporation of Pyl into a nascent protein in response to an amber codon during translation

In contrast to the incorporation of selenocysteine, no specific mRNA secondary structure is required for Pyl incorporation.⁴⁸⁻⁵¹

4.2 Incorporation of NCAAs into proteins

The incorporation of non-canonical amino acids (NCAAs) into proteins can add various new functionalities to the standard amino acid set. Examples for possible applications are given in section 4.2.1. Possible ways of introducing unnatural amino acids into proteins are described in the sections 4.2.2.1 and 4.2.2.2.

4.2.1 Applications of unnatural amino acids in proteins

4.2.1.1 Incorporation of new functionalities

Unnatural amino acids can, without the need of further modification, directly introduce useful moieties into proteins.⁵² These moieties can e.g. simulate natural post translational modifications (PTMs) to enable the elucidation of their biological function which has been shown successfully for lysine methylations, acetylations, acylations or phosphorylations.⁵³⁻⁵⁹ The directly introduced functionalities are of course not limited to naturally occurring PTMs. Examples of NCAAs with new functionalities that have been introduced into proteins include also fluorescent, metal-chelating, redox-active, photo-active, spin-labeling, stability increasing and backbone-extending properties.⁶⁰⁻⁶⁶ Novel NCAAs enable scientists e.g. to crosslink the protein of interest or create electron transfer pairs to study its interaction with binding partners.⁶⁷⁻⁷⁰ Conformational changes in the protein structure can as well be analyzed e.g. by specifically introduced vibrational probes.⁷¹ Another application example is the specific design of artificial protein multimers which was facilitated by the introduction of metal-chelating NCAAs.⁷² Even the control of protein activity using a small molecule switch inside live cells was shown to be possible using a specifically designed NCAA.⁷³

The most versatile group of NCAAs introduces chemical handles, which can be specifically modified after protein expression allowing the fusion of biology and chemistry.

4.2.1.2 Protein conjugation

Examples of pharmaceutically applied protein conjugates

Adding new functionalities to various biomolecules represent an area of extensive interest in the field of chemical biology. There are many reasons for adding new properties to

biomolecules via covalent attachment of small synthetic molecules. For example, such conjugates can facilitate investigations of protein function and structure, be used in cell labeling studies, generate novel biomaterials or be useful in biomedical applications.^{68,74-77}

Protein conjugates already applied in biopharmaceuticals are e.g. conjugates of PEG and small proteins (below 70 kDa), which increase the half-life of the drug by increasing their size and preventing rapid renal clearance.^{78,79} The frequency of drug-injections can be decreased dramatically as e.g. in the case of N-terminally PEGylated granulocyte-colony stimulating factor (G-CSF, INN: *Pegfilgrastim*). The number of injections of this drug which stimulates the production of granulocytes and is frequently applied during chemotherapy can be reduced from a daily administration (*Filgrastim*) during therapy to a once-per-cycle treatment (*Pegfilgrastim*).^{80,81}

Another example for pharmaceutically relevant protein conjugates are antibody-drug-conjugates (ADCs). This class of biopharmaceuticals contains highly potent toxins fused to malignant cell-targeting antibodies which deliver their poisonous payload directly into the target cell.⁸²⁻⁸⁵ The ADC *Trastuzumab emtansine* for example directs the cytotoxic agent emtansine to HER2 positive breast cancer cells which destroys them by binding to tubulin.⁸⁶⁻⁸⁸

Protein conjugation using natural amino acids

One of the major goals in the production of protein conjugates is the generation of a homogenous product, preferably consisting of only one conjugation species. Principally, various methods exist for the attachment of synthetic molecules to proteins. Most applied methodologies used in commercialized biopharmaceuticals exploit functional groups present in the 20 proteinogenic amino acids. Among these, the ϵ -amino group of lysine, the thiol group of cysteine, or the N- or C-terminal groups are most commonly applied.^{75,89,90} If the addressed functional group is present several times on the surface of the protein of interest, as is usually the case for lysine amino groups, the synthetic molecule is added multiple times and non-site-specifically. The modification of the lower abundant cysteines, however, bears the danger of disrupting naturally occurring disulfide bonds and therefore influences the secondary structure of the protein. Depending on the protein, N-

terminal attachments were shown to be relatively selective, even though cross reactions with ϵ -amino groups of lysines are commonly observed. Additionally, again depending on the protein, terminal conjugations were shown to potentially influence the activity of the biomolecule.⁸⁹

In order to increase the site specificity of the conjugation reaction, peptide tags with specific amino acid sequences can be used. Two examples of repeating amino acid tags are the tetracysteine tag which directly reacts with biarsenic compounds and the hexahistidine-tag which binds Ni(2+)-nitrilotriacetate (NTA) probes.^{91,92} Another family of peptide/protein tags involves enzymes which specifically recognize and modify the fused peptide. Two examples of this growing family have to be mentioned. The SNAP and Clip tag are engineered transferases which transfer small molecules onto themselves and enable direct small molecule attachment and the aldehyde tag.^{93,94} The latter contains a cystein-thiol which is transformed into an aldehyde by FGE, the formylglycine-generating enzyme.⁹⁵ The aldehyde can afterwards be specifically modified by hydroxyl amines or hydrazines.

These peptide tags are easy to introduce and versatile tools for many applications. Nevertheless, they are big compared to single amino acids and usually introduced on one terminus which may disrupt the structure and function of the protein.

Protein conjugation using NCAAs

An uncompromising chemo- and site-specific protein conjugation technique takes advantage of unnatural, non-canonical amino acids (UAAs/NCAAs). These NCAAs can be incorporated into proteins by using various strategies. The first successful attempts to introduce unnatural amino acids into peptides were made by solid phase peptide synthesis, which is limited to shorter peptides consisting of up to around 100 amino acids.^{96,97} Semisynthetic strategies allow the combination of NCAA containing synthetic peptides with purified proteins by native chemical ligation.⁹⁸⁻¹⁰¹ These (semi-)synthetic strategies result in specific products but are tedious especially for the generation of bigger proteins.

Another promising approach for the incorporation of NCAs utilizes the natural translational machinery. The NCAs are charged onto tRNAs which are recognized by the ribosome and incorporated in response to the respective mRNA-codon. This idea was first applied in *in vitro* translation systems.¹⁰² Transferring the idea of genetically encoded NCA-incorporation to *in vivo* systems represents the major breakthrough in the field. This methodology is described in more detail in the following sections.

4.2.2 Site-specific, genetically encoded incorporation of NCAs

The genetic code does not contain blank codons that could be reassigned to new sense codons. Therefore, smart approaches had to be developed to fulfill the task of ribosomal and site specific incorporation of NCAs into proteins.¹⁰³ So far, there are three successful strategies for this task: Rare codon reassignment, amber stop codon suppression and the usage of quadruplet codons. All of these strategies share the introduction of an orthogonal tRNA/aminoacyl-tRNA synthetase (tRNA_{YYY(Y)}/aaRS) pair into a host cell (see Figure 4.7). The tRNA is charged with the NCA by the aaRS. Later on, the ribosome uses the NCA-charged tRNA to specifically translate the assigned complementary mRNA codon. Therefore, the strategies mainly differ in the assigned mRNA codon, at which the NCA is incorporated into the protein of interest.

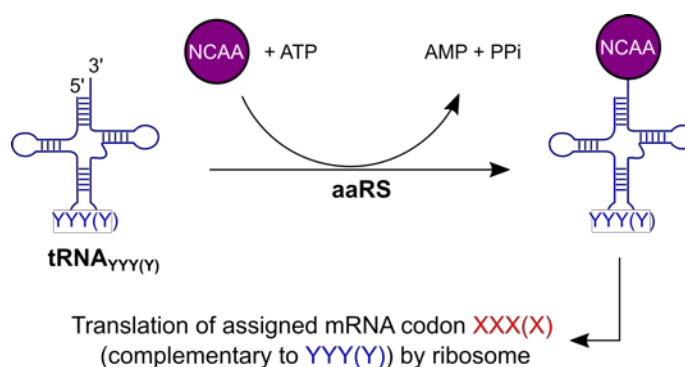


Figure 4.7 Principle of site-specific, genetically encoded incorporation of NCAs. Depending on the assigned mRNA codon (red) strategies can be divided into rare codon reassignment, amber stop codon reassignment or assignment of quadruplet codons.

In rare codon reassignment, an uncommon natural amino acid codon as e.g. AGG, AGA (both arginine) or AUA (isoleucine) in *E. coli* is recoded to a NCA by using orthogonal tRNA/aminoacyl-tRNA synthetase pairs.¹⁰⁴⁻¹⁰⁷ Principle and challenges are similar to the recoding of amber stop codons (described later).

Instead of reprogramming an existing triplet-codon and disturbing its natural function, the use of artificial quadruplet codons allows, in theory, the generation of $4^4 = 256$ new sense codons. In this technology, the genetic information of special tRNAs which contain four bases in the anticodon loop (e.g. tRNA_{AGAG}) are introduced together with the respective aminoacyl synthetase gene into the genome of the host organism.¹⁰⁸ Therefore, a

quadruplet codon in the mRNA (e.g. CUCU) does either result in a frame shift, or the translation to the NCAA (frame shift suppression). The principle could already be shown in *E. coli*, *Xenopus oocytes* and mammalian cells but is currently not competitive compared to triplet codon reassignment strategies.¹⁰⁹⁻¹¹⁴

Amber codon suppression is, so far, the most versatile and successful approach used for *in vivo* NCAA incorporation. This methodology was also the key technology of this thesis and is described in more detail in the following sections.

4.2.2.1 Amber suppression via orthogonal tRNA/aminoacyl-tRNA synthetase pairs

Three of the 64 triplet codons lead to the termination of ribosomal protein synthesis by recruiting release factors (e.g. RF1 or RF2 in *E. coli*) which results in the disassembly of the ribosome subunits. As described above, these stop codons are called ochre (UAA), opal (UGA) and amber (UAG). In *E. coli*, only 273 (7%) of all genes are terminated by the amber codon and essential genes are only in very few cases terminated by this stop codon.¹¹⁵⁻¹¹⁷

Several species do not employ the amber codon as a stop codon, but as a sense codon for a canonical amino acid. One example is *Methanococcus jannaschii*, a thermophilic methanogenic Archean, which uses the amber codon as a sense codon for tyrosine.¹¹⁸ The responsible tRNA / aminoacyl tRNA-synthetase pairs (*Mj*TyrRS/*Mj*tRNA^{Tyr}) have been transferred to standard laboratory strains like *E. coli* or CHO cells and successfully used for the incorporation of NCAs. Recent studies employed other aaRS/tRNA pairs as e.g. the *E. coli* pair of tryptophane.^{52,119}

To accomplish the incorporation of NCAs, Schultz et al. developed a general approach which includes a series of selection steps.¹²⁰ The major challenge here was to transform *Methanococcus jannaschii* tRNA and aminoacyl tRNA synthetase into an orthogonal pair by directed evolution. Orthogonality in this context means, that the tRNA exclusively recognizes amber codons and the synthetase only charges it with the intended NCAA. Therefore, a cross reaction with the endogenous *E. coli* expression system can be excluded. In brief, the directed evolution consists of two major steps. First, a tRNA library is created and suitable candidates which are not accepted by endogenous *E. coli* aminoacyl tRNA synthetases are selected. Second, an aminoacyl tRNA synthetase library is created and screened for candidates which are able to charge this tRNA with the desired amino acid.

For the selection process, Schultz *et al.* introduced the amber codon either into the toxic barnase gene for negative selection or into the β -lactamase ampicillin resistance gene for positive selection. The first genetically incorporated NCAAs were the synthetic amino acid O-methyl-L-tyrosine (**1**), which was incorporated using the described evolved *Methanococcus jannaschii* tRNA and aminoacyl tRNA synthetase pair.¹²⁰ This tyrosine derivative served as a proof-of-concept and was followed by more than 50 NCAAs up to now which were introduced into various proteins using this system.^{60,121} Figure 4.8 illustrates the variety of incorporated amino acids by showing some of the most remarkable examples. The ketone-bearing *p*- and *m*- acetylphenylalanines (**2**, **3**) were the first ribosomally incorporated NCAAs, enabling the bio-orthogonal modification of proteins by reaction with hydroxylamines.^{122,123} Especially *p*-acetylphenylalanine (**2**) was chosen for a variety of applications and was one of the first site-specifically incorporated NCAAs in eukaryotic and mammalian cells.¹²⁴⁻¹²⁶ Also other chemically interesting functionalities like e.g. β -diketone (**4**), alkene (**5**), alkyne (**6**) or azide (**7**) moieties were introduced using this technique.¹²⁷⁻¹³¹ The alkyne and azide tyrosyl variants are able to react in copper mediated click reactions (see also section 4.3.4) and are therefore of special interest. Other impressive examples are the redox active 3,4-dihydroxy phenylalanine (**8**), the ferrocene containing cysteine variant **9**, the bidentate metal chelating bipyridylalanine **10**, the popular photo-reactive *p*-benzoylphenylalanine **11** and the photoisomerizable fluorescent phenylalanine-4'-azobenzene **12**.¹³²⁻¹³⁸

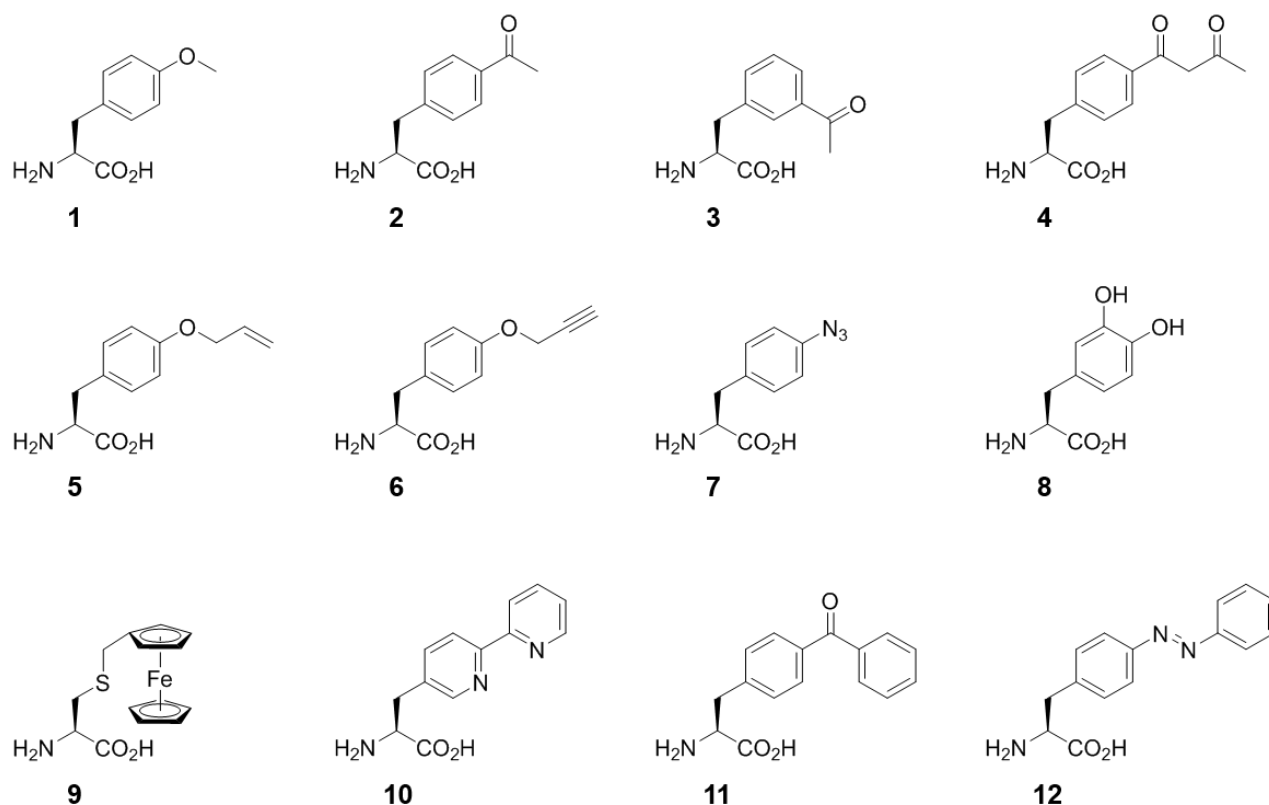


Figure 4.8 Examples of incorporated NCAAs using evolved *Methanococcus jannaschii* tyrosyl tRNA and tRNA aminoacyl synthetase pairs

These examples impressively illustrate the diversity of functional groups that have been incorporated using the *Mj*TyrRS/*Mj*tRNA^{Tyr} pair but also show the structural limitations of this system. Even though the *M. jannaschii* tyrosyl tRNA synthetase has been evolved extensively, the incorporated NCAAs all show high similarity to tyrosine. Therefore, other aaRS/tRNA pairs have been used as starting points for the creation of orthogonal pairs in *E. coli* such as Asp, Gln, Tyr or Phe aaRS/tRNA pairs from *Saccharomyces cerevisiae* or hybrid pairs where the aaRS and the tRNA originate from different organisms.^{109,139-144}

All of these artificial systems have in common that especially bulkier moieties e.g. interesting bio-orthogonal reaction targets like norbornenes or strained alkenes and alkynes could not be incorporated due to structural restrictions in the aaRS active site. Therefore, amongst other interesting properties, the less restrictive pyrrolysine PylRS/tRNA^{Pyl} pair (see sections 4.1.3.3 and 4.2.2.2) was shown to be an interesting alternative and a useful addition for site-specific amino acid incorporations by amber suppression.

4.2.2.2 Amber suppression via the pyrrolysine system

The discovery of the genetic incorporation of pyrrolysine in response to the amber codon immediately added a promising aaRS/tRNA pair to the already existing amber suppression tools. In its natural context in *Methanosarcinae* strains, but also in bacteria like *Desulfitobacterium hafniense*, the PylRS/tRNA^{Pyl}_{CUA} pair shows no cross reactivity with the endogenous protein synthesis machinery. Therefore, it was believed to show natural orthogonality without the need of further adaptations even in laboratory strains like *E. coli*, *S. cerevisiae*, *Caenorhabditis elegans*, *Drosophila melanogaster* or mammalian cells and has proved true.^{42,51,145,146} Even the transfer in live mouse brain cells was shown to be possible.¹⁴⁷ PylRS/tRNA^{Pyl}_{CUA} pairs of *M. mazei*, *M. barkeri* and *D. hafniense* have been successfully transferred into different species without losing orthogonality.^{145,148-150} Figure 4.9 depicts the incorporation of NCAAs via the pyrrolysine system in host cells.

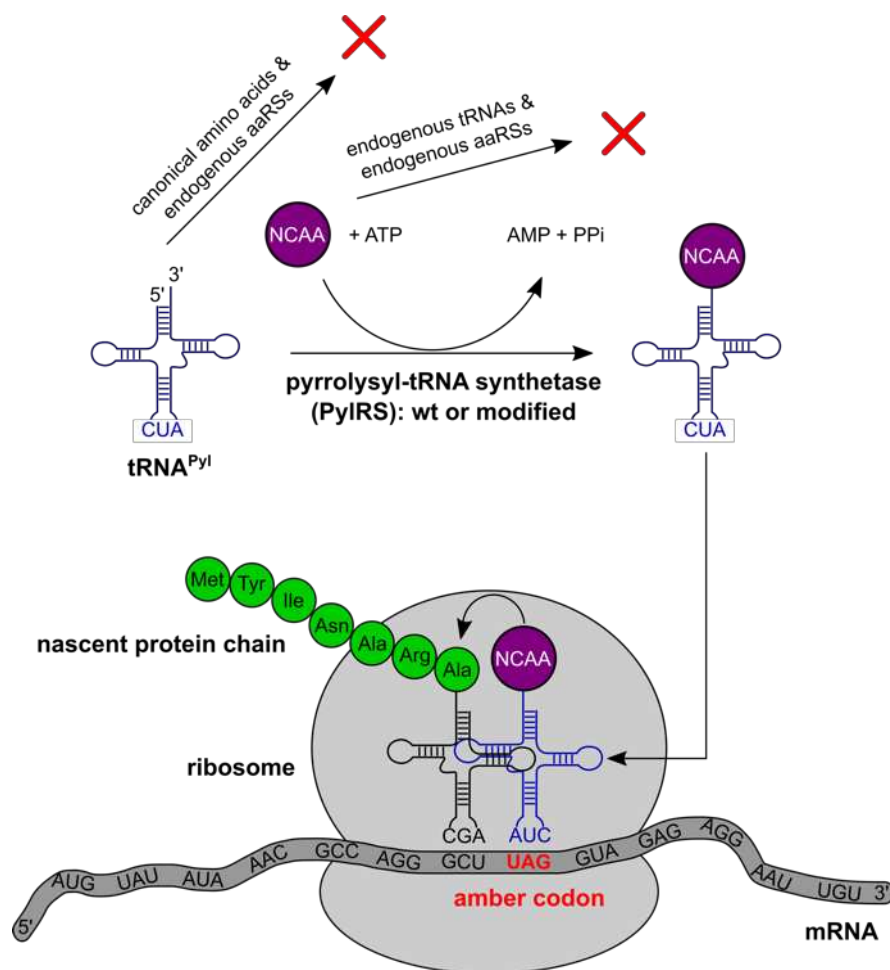


Figure 4.9 Incorporation of NCAAs by introducing the pyrrolysine system into host cells. Orthogonality is symbolized by red x. Depending on the NCA, wt or an adapted PylRS variant has to be introduced (see sections below).

4.2.2.2.1 NCA incorporation via the wt PylRS/tRNA^{Pyl}_{CUA} pair

Besides the natural amino acid pyrrolysine, more than 20 analogues could be incorporated by transferring the wt PylRS/tRNA^{Pyl} pair to host cells like *E. coli* without the need for further modification of the system.^{53,148,150-163} **Figure 4.10** shows selected Pyl analogs which are accepted by unmodified PylRS and illustrate the promiscuity of wt PylRS. Derivatives **13** and **14** served as the first artificial substrates for studying PylRS since the total chemical synthesis of Pyl is tedious.^{151,164} The Boc-protected methyllysine **15** can be deprotected after site-specific incorporation to yield N-ε-methyllysine, a common posttranslational modification (PTM). This allows the position-specific function determination of this regulatory element e.g. in histones.^{53,162} Lysine derivatives **16** - **21** carry handles for bio-orthogonal conjugation reactions which will be described in section 4.3.^{158-160,163,165} The

target protein expression yield strongly depends on the acceptance of the desired Pyl-analog by the transferred Pyl-translation machinery. Amino acids **19** - **21** for instance can be incorporated using the *wt* PylRS/tRNA^{Pyl} pair but the obtained yields can be improved by adaptations of the system which will be described in the following sections.

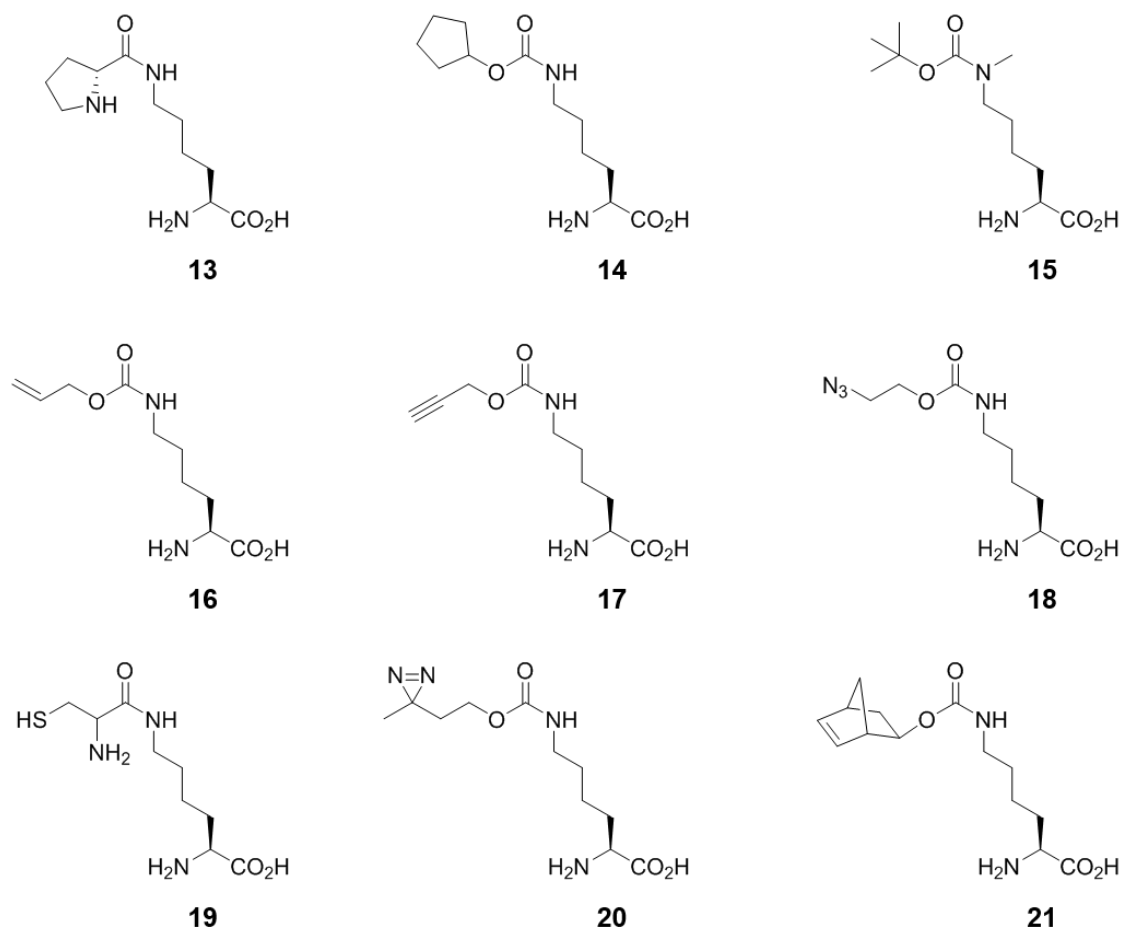


Figure 4.10 Examples of incorporated NCAAs using *wt* tRNA^{Pyl}/PylRS pairs

4.2.2.2.2 PylRS structure

The key element for efficient incorporation of bulky NCAsAs with high structural diversity is PylRS since it recognizes Pyl-analogues and loads them onto tRNA^{Pyl} (see section 4.1.3.3). Besides the low binding selectivity towards the tRNA anticodon, elucidation of the PylRS structure from *M. mazei* revealed a deep hydrophobic pocket as the main recognition site for Pyl.^{51,149,150,166-168} This pocket has high affinity for the pyrroline moiety of Pyl ($K_d = 53 \mu\text{M}$) but binds it mainly by relatively nonspecific hydrophobic interactions (see Figure 4.II). Additionally, PylRS does not contain an editing domain which could hydrolyze misacylated tRNAs like many other aaRSs do.¹⁶⁹⁻¹⁷¹ Both findings explain the very unique substrate promiscuity for Pyl-related hydrophobic lysine variants.

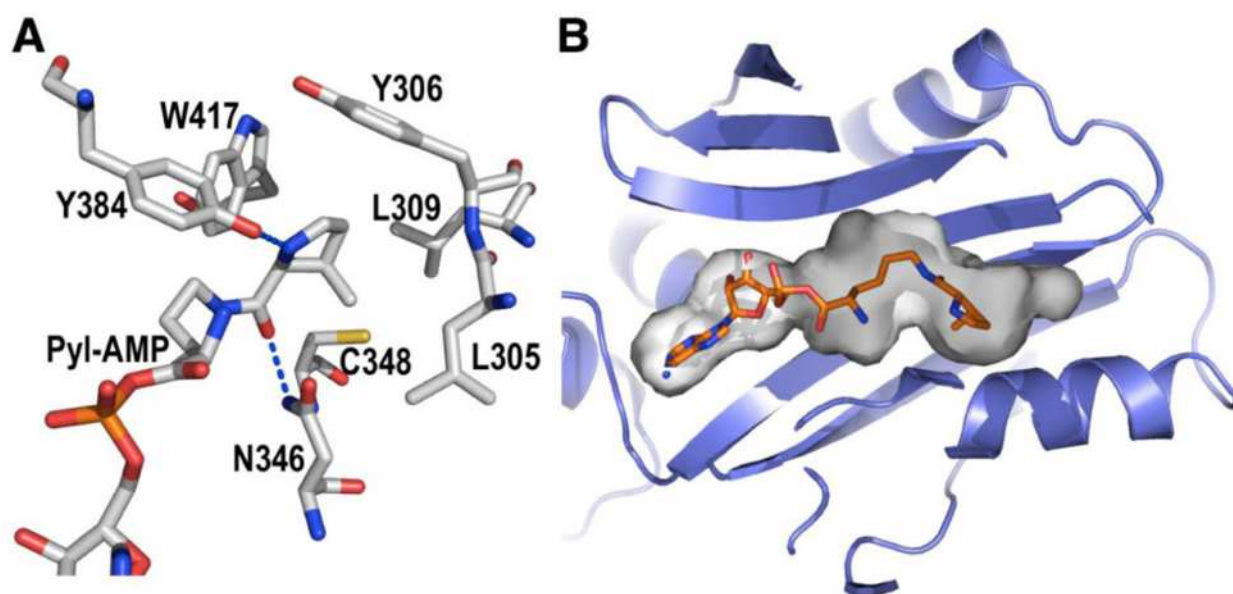


Figure 4.II Structural details of *M. mazei* PylRS. A: Recognition site of Pyl-AMP. B: Substrate binding pocket of PylRS harboring Pyl-AMP.

4.2.2.2.3 NCAA incorporation via engineered PylRS/tRNA^{Pyl}_{CUA} pairs

The detailed structural knowledge of PylRS paved the way for directed evolution of the recognition pocket similar to what was done before with *Mj*TyrRS (section 4.2.2.1). Conveniently, modifications of the recognition site of PylRS have no noticeable effect on the actual activity of the enzyme since it is located relatively far away from the active site where the acylation of the amino acid takes place. Research groups who worked on the rational design or directed evolution of the recognition site independently identified five major residues of PylRS with high influence on the accepted Pyl analogs. In case of *M. mazei* PylRS these residues are Y306, L309, N346, C348 and Y384 (Figure 4.11A). An exchange of Y306 to a smaller amino acid like alanine or glycine for example enlarges the hydrophobic cavity and allows bulkier moieties to enter. Another example is the exchange of Y384 by a phenylalanine, which eliminates an hydroxy group and therefore results in even higher hydrophobicity of the recognition cavity and greater promiscuity of the enzyme. Recently, phage assisted continuous evolution (PACE), a novel directed evolution technique, was applied to PylRS, which increased the achieved NCAA containing protein yields.¹⁷²

At least 30 different mutants of PylRS were developed which have led to an impressively diverse selection of more than 80 incorporated NCAs (examples are shown in Figure 4.12).^{1,53-55,148,160-162,173-197} The engineered PylRS mutants show either higher incorporation activities than their *wt* equivalents or are required to make the incorporation of the desired NCAs possible because the *wt* PylRS do not show any detectable activity on them.

Amino acids incorporated by codon reassignment include posttranslationally modified lysines like acetyllysine **22**, or the photocaged N-ε-methyllysine **23**.^{54,187}

The genetical incorporation of NCAA **22** greatly facilitated site specific functional studies of acetyllysine, the most abundant acyl-PTM in mammalian cells e.g. in the nucleosome or p53.^{55,198-201}

Furthermore, evolved PylRSs enabled the genetical incorporation of novel, bulkier conjugation groups like alkene **24**, azide **25**, protected native chemical ligation handle **26** or the furan containing protein-RNA crosslinker **27**. Especially the strained trans-

cyclooctene- (**28** and **29**), norbornene- (**21** and **30**) or cyclooctyne- (**31** and **32**) carrying lysine derivatives made unprecedented ultra-rapid protein conjugations possible (see section 4.3).^{1,174-177}

Another example is the first incorporated radical NCAA **33** which contains a spin label and has been used to report intramolecular distance distributions of two sites in a model protein *in vivo* by double-electron electron resonance measurements.¹⁹²

A whole new family of phenylalanine derivatives including phenylalanine itself (**34**), *m*-azide-phenylalanine **35** or terminal alkyne containing tyrosine derivative **36** was recently added to the set of NCAAs incorporated by the Pyl system.^{194,195} These non-lysine derivatives were tested, since bacterial PheRS shows high structural similarities to PylRS.^{193,197,202-204} The mutagenesis of PylRS towards acceptance of Phe analogs was more successful than originally anticipated and it could be shown that the introduction of only two PylRS-mutations (N346A and C348A) is sufficient for the incorporation of more than 40 Phe analogs.^{194-196,205} Although most of these phenylalanine derivatives can also be incorporated with the described MjTyrRS/tRNA^{Tyl} pair (see section 4.2.2.1), their acceptance by evolved PylRS underlines the potential of this versatile genetic code expansion tool.

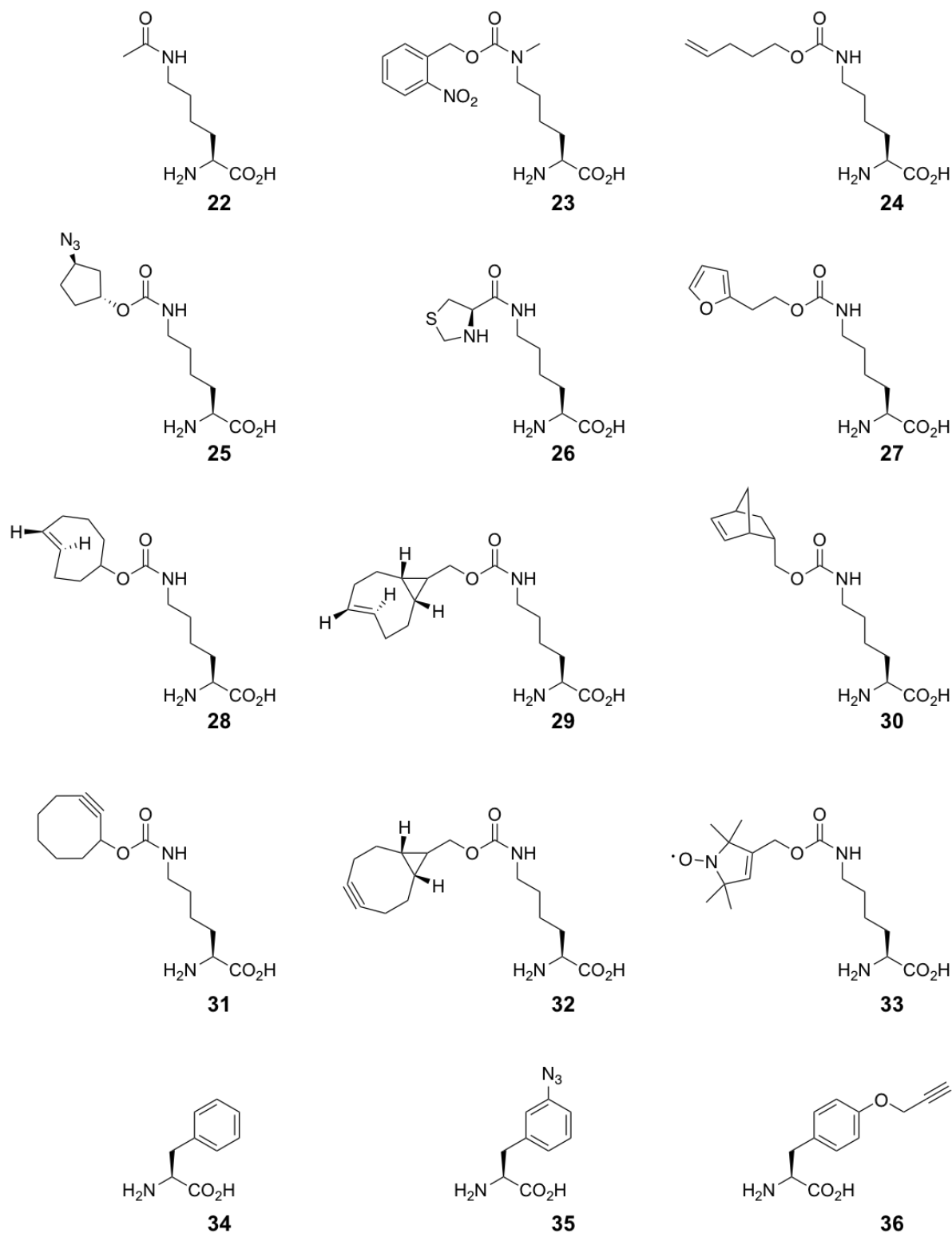


Figure 4.12 Examples of incorporated NCAAs using evolved *Pyl*RSs with wild type *tRNA*^{*Pyl*}

4.3 Bioorthogonal modifications of unnatural pyrrolysine homologs

4.3.1 Bioorthogonality and practical aspects of protein labeling strategies

The ability to label NCAs site specifically and bioorthogonally opens new fields in protein sciences. The two major applications are the *in vitro* labeling of purified proteins for studying or extending their properties and the *in vivo* labeling of selected proteins e.g. for visualization of their distribution on the surface or inside a living cell.

Depending on the purpose, different characteristics of the labeling chemistry become essential. When labeling a purified protein, the main focus lies on the generation of a specific modification product in high yield and therefore low amounts of side product. The kinetics of the reaction are less important in this case since the conjugation can be carried out usually with high excess of the modification reagent.²⁰⁶ In contrast, labeling of proteins in cells necessarily requires rapid reaction kinetics ($k_2 \geq 1 \text{ M}^{-1} \text{ s}^{-1}$) because the probe must be applied in small amounts to prevent its possible toxicity and to exclude interference with the living system. In addition, the ligation reaction must efficiently proceed before the reagent gets metabolized or is cleared from the cell.²⁰⁷⁻²¹¹ The labeling yield and homogeneity of the labeling product are usually less important in this case.

Independently of the protein conjugation purpose, the most important property of protein labeling chemistry is bioorthogonality. This term sums up several characteristics and its evaluation defines whether a chemical reaction can be used for biomolecule labeling.^{212,213} First of all, to be considered bioorthogonal, a reaction has to be applicable under biological conditions. This implies reaction conditions in aqueous solutions, pH values between 4 and 10 and temperatures of maximum 40 °C. Secondly, the reaction has to be chemically and biologically inert which means, that the formed linkage has to be stable and none of the reaction reactants or products should unintentionally interfere with the native functionality of the protein or the studied organism. Furthermore, the reaction has to be selective only for the unnaturally incorporated functional group and not side react with other biological groups. In the case of purified protein conjugations, these competing groups only include the 20 proteinogenic amino acids, whereas in *in vivo* labeling also other functional groups e.g. aldehydes or alkenes in fatty acids are present which makes

selectivity more challenging. Therefore, a differentiation between truly bioorthogonal and chemoselective reactions can be made.^{212,213}

Finally, a rather practical aspect in the identification of the optimal protein labeling strategy and chemistry is the accessibility of the reagents. This includes the already described incorporation efficiency of the NCAA but also the chemical-synthetic effort that has to be made to obtain both the NCAA and the labeling reagent. Especially the NCAA, which is added in huge excess to the expression medium, should be easily accessible via high yielding synthetic routes.

Unfortunately, a generally favorable bioorthogonal modification strategy does not exist since all of the described aspects have to be considered and evaluated according to the labeling purpose. Therefore, a set of different bioorthogonal reaction methods with their individual advantages and disadvantages has been developed. In the following subsections, the most widely used modification methods, which have been successfully applied on NCAs introduced by the Pyl incorporation system will be discussed.

4.3.2 Ketone-hydrazide/hydroxylamine reactions

Aldehydes and ketones are small, synthetically accessible and versatile functional groups which have been used extensively as handles for biomolecule.²¹⁴⁻²¹⁷ While aldehydes were considered less suited for protein modification reactions due to possible amine additions, the less reactive ketones have been successfully introduced and modified in proteins. In fact, the first modification handles introduced in proteins were ketone containing amino acids and have since then been successfully applied in various protein modification studies.⁶⁰

Ketones can be modified under slightly acidic conditions (pH 4-6) by nucleophiles with strong α -effect like hydrazines or hydroxylamines to form hydrazones and oximes, respectively (Figure 4.13). The reaction can be catalyzed by the addition of aniline which activates the keto-group by formation of an intermediate Schiff-base and enables the use of neutral reaction conditions.²¹⁸⁻²²⁰ The catalyzed reaction reaches second order rate constants of up to $170 \text{ M}^{-1} \text{ s}^{-1}$.²¹⁸

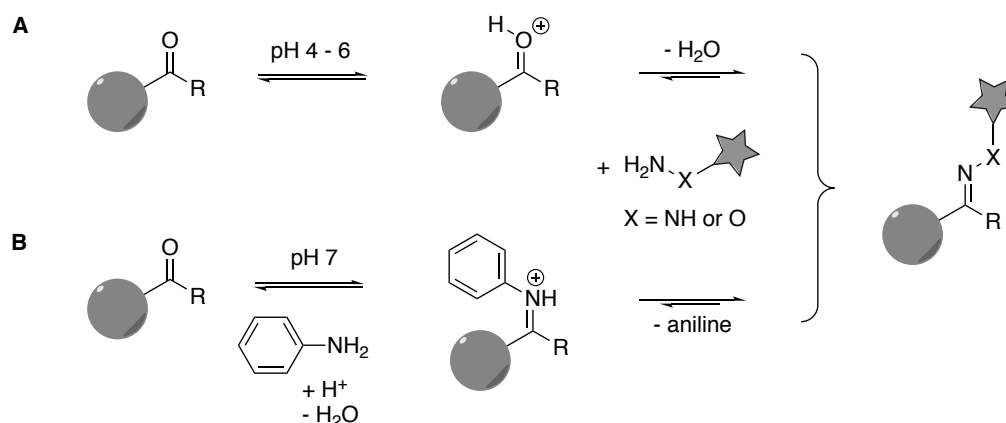


Figure 4.13 Reaction of aldehydes and ketones with hydrazines or hydroxylamines. A: Acid catalyzed, B: Aniline catalyzed under neutral conditions

Owing to the small size of these functional groups but rather slow kinetics (second order rate constants between 10^{-4} and $10^{-3} \text{ M}^{-1}\text{s}^{-1}$ for uncatalyzed reaction) or the need of a catalyst, the reaction is perfectly suited for *in vitro* protein labeling but unfavorable for live-cell protein labeling because large excesses of labeling agents are needed.^{221,222}

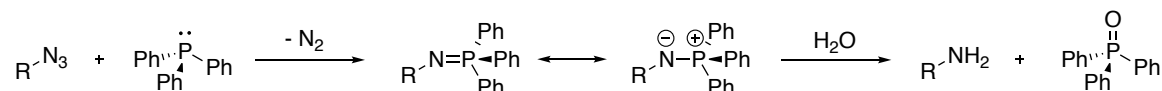
4.3.3 Staudinger ligations

One of the first bioorthogonal reactions developed utilizes azides as the reactive chemical group. The small size, absence in nature, synthetic accessibility and inertness toward natural functional groups make azides highly versatile and useful reagents in bioconjugations.^{3,60,148,206,223-232} Azides have been successfully incorporated into various biomolecules which include glycans, DNA, RNA, fatty acids and also proteins.²³³⁻²³⁷

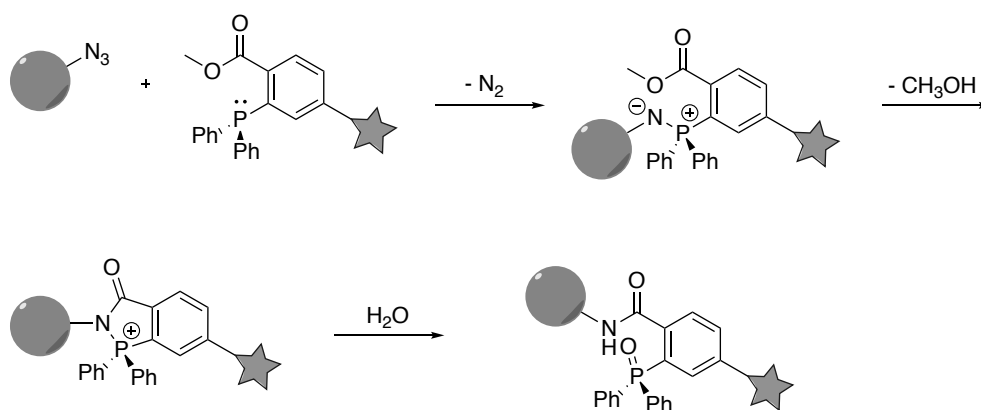
In analogy to the well-established Staudinger reduction (Figure 4.14 A), the group of C. Bertozzi developed in 2000 the so-called Staudinger ligation reaction.²³⁸⁻²⁴⁰ The original reduction reaction, first described in 1919, starts with the nucleophilic attack of triphenylphosphine and the formation of an Aza-ylide after loss of nitrogen from the azide. The Aza-ylide can afterwards be hydrolyzed to yield a primary amine as the reduction product and triphenylphosphine oxide as byproduct.²⁴¹ In the Staudinger ligation reaction an internal ester group ortho to the phosphorous atom is introduced as electrophilic trap, which leads to intramolecular ring formation. After hydrolysis an amide bond is formed and connects the biomolecule with the intended substituted phosphine oxide (Figure 4.14 B).^{240,242}

Fusion of the intended label directly onto the carbonyl of the ester group even makes traceless Staudinger ligations possible which only result in an amide bond between the biomolecule and the marker (Figure 4.14 C).²⁴³

A: Staudinger reduction



B: Staudinger ligation



C: Traceless Staudinger ligation

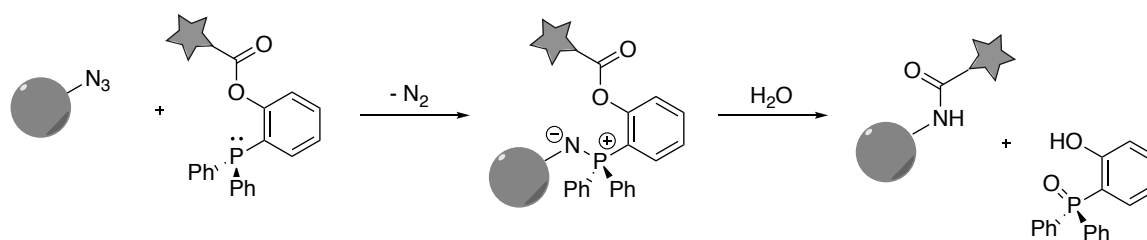


Figure 4.14 A: Staudinger reduction B,C: Staudinger ligations

Due to the rather slow reaction kinetics with second order rates of around $10^{-3} \text{ M}^{-1}\text{s}^{-1}$ and the oxidation sensitivity of phosphines, high concentrations of labeling reagent are required.²⁴¹ Staudinger ligations are therefore mainly used for e.g. *in vitro* labeling of purified proteins or *in vivo* labeling of highly abundant biomolecules.

4.3.4 Azide-alkyne reactions

In 1963, Rolf Huisgen studied the reaction of azides with terminal alkynes to yield stable triazoles.²⁴⁴ The reaction is interesting as bioorthogonal tool since it involves only small functional groups without cross reactivity to biomolecules. Although the reaction is highly exergonic ($\Delta G^\circ = -60$ kcal/mol), it requires elevated temperature and pressure due to a high activation barrier which makes it not biocompatible in its original form.²⁴⁵ 39 years later, Sharpless et al. and Meldal et al. reported independently, that copper(I) salts are able to catalyze the Huisgen reaction in water and increase the second order rate constant k_2 by six to seven orders of magnitude.^{246,247} Moreover, the addition of Cu(I) was shown to influence the regioselectivity of the reaction yielding only the 1,4-substituted triazoles.^{248,249} The initially proposed reaction mechanism of the Cu(I) catalyzed azide-alkyne cycloaddition (CuAAC) including one catalytic copper ion forming a six membered ring intermediate was revised by in-depth isotopic studies.²⁵⁰ The key difference found through these studies is the formation of a dinuclear copper intermediate as shown in Figure 4.15 B.

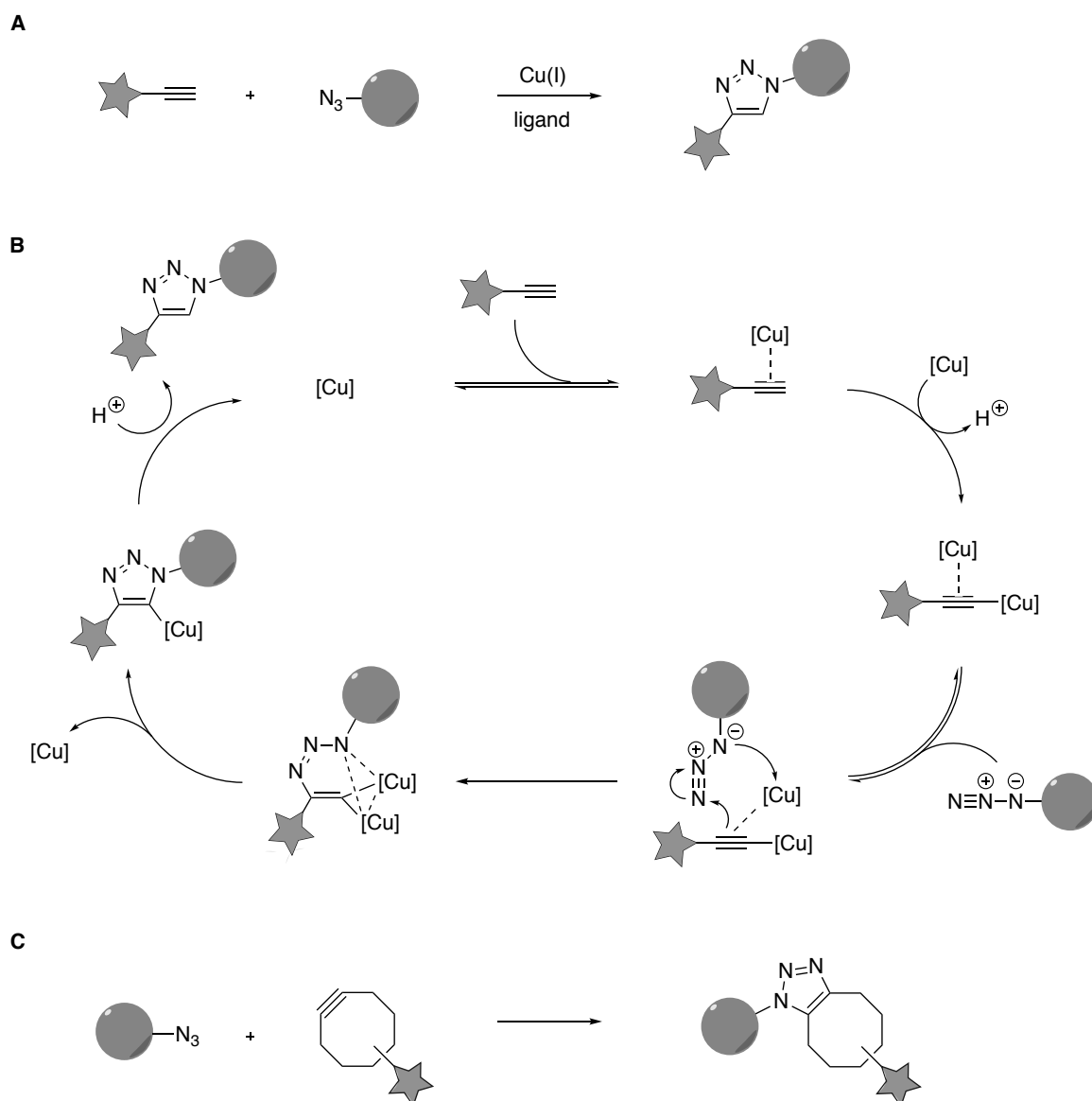


Figure 4.15 Azide-alkyne reactions

Sharpless et al. used CuAAC as the prime example for "click chemistry", a biology-inspired concept of modular synthetic chemistry.²⁵¹ CuAAC has also been proven very useful as biomolecule modification reaction in many biological studies including nucleic acid, protein or lipid functionalization.^{60,131,231,237,252-256} The disadvantage of this reaction is the need of catalyzing copper salts which have been proven to have potentially toxic effects, shown on e.g. *E. coli* and mammalian cells.²⁵⁶⁻²⁶⁰ The toxic effect of Cu(I) can be reduced by the use of specific water soluble ligands which form activated complexes and therefore reduce the absolute amount of copper needed and further lead to acceleration of the reaction by stabilizing the active catalytic Cu(I) species.²⁶¹⁻²⁶³ Even applications in living organisms like zebrafish have been shown.²⁶⁴⁻²⁶⁶ Ligand-assisted CuAAC reactions

typically reach second order rate constants of 10-200 M⁻¹s⁻¹ but need copper concentrations of 100 - 500 µM and therefore have limitations when labeling biomolecules in living cells.²⁶⁷⁻²⁶⁹ CuAAC reactions with lower copper concentrations down to 10 µM can be achieved by copper-chelating azides such as picolyl azides.²⁷⁰

An azide-alkyne reaction without the need of a catalyst takes advantage of the intrinsic reactivity of ring strained alkynes (Figure 4.15 C). The reaction of cyclooctynes, the smallest stable and isolable cycloalkyne, has been described already more than 70 years ago before it was first applied in bioorthogonal labeling reactions.²⁷¹⁻²⁷³ Strain-promoted-alkyne-azide cycloadditions (SPAAC) have been successfully applied in a series of biomolecule labeling studies including site-specific protein modifications.^{2,3,225} The reaction is highly selective and non-toxic but shows rather low reaction kinetics with k_2 ranging between 0.0012 M⁻¹s⁻¹ for unmodified cyclooctynes and 1 M⁻¹s⁻¹ for optimized cyclooctynes including e.g. difluorocyclooctynes (DIFO) or dibenzocyclooctynes (DIBO).^{274,275}

4.3.5 Inverse electron-demand Diels-Alder cycloadditions

The cascade of an inverse electron demand Diels-Alder cycloaddition (iEDDA) followed by a *retro*-Diels-Alder (rDA) reaction between 1,2,4,5-tetrazines and strained alkenes like norbornenes or *trans*-cyclooctenes has been identified as a powerful and rapid ligation reaction in synthetic chemistry.²⁷⁶⁻²⁷⁸ The cycloaddition step of the reaction sequence is termed inverse electron demanding since, in the transition state of the reaction, the LUMO of the electron-poor tetrazine interacts with the HOMO of the electron rich dienophile. In classic Diels-Alder reactions the situation is opposite, the dienophile LUMO interacts with the HOMO of the diene.²⁷⁹ The iEDDA results in a strained bicyclic intermediate, which rapidly releases nitrogen in a subsequent rDA reaction (Figure 4.16 A). In protic solvents, the resulting 4,5-dihydropyridazine isomerizes to the final product 1,4-dihydropyridazine.²⁸⁰

The iEDDA reaction was shown to work under aqueous conditions for the first time by Hildebrand et al. using norbornene and Fox et al. using *trans*-cyclooctenes as dienophiles.^{281,282} (Figure 4.16B and C) The finding that aryl-substituted tetrazines are sufficiently stable in water was the key discovery for application of this chemistry on biomolecules.^{281,283}

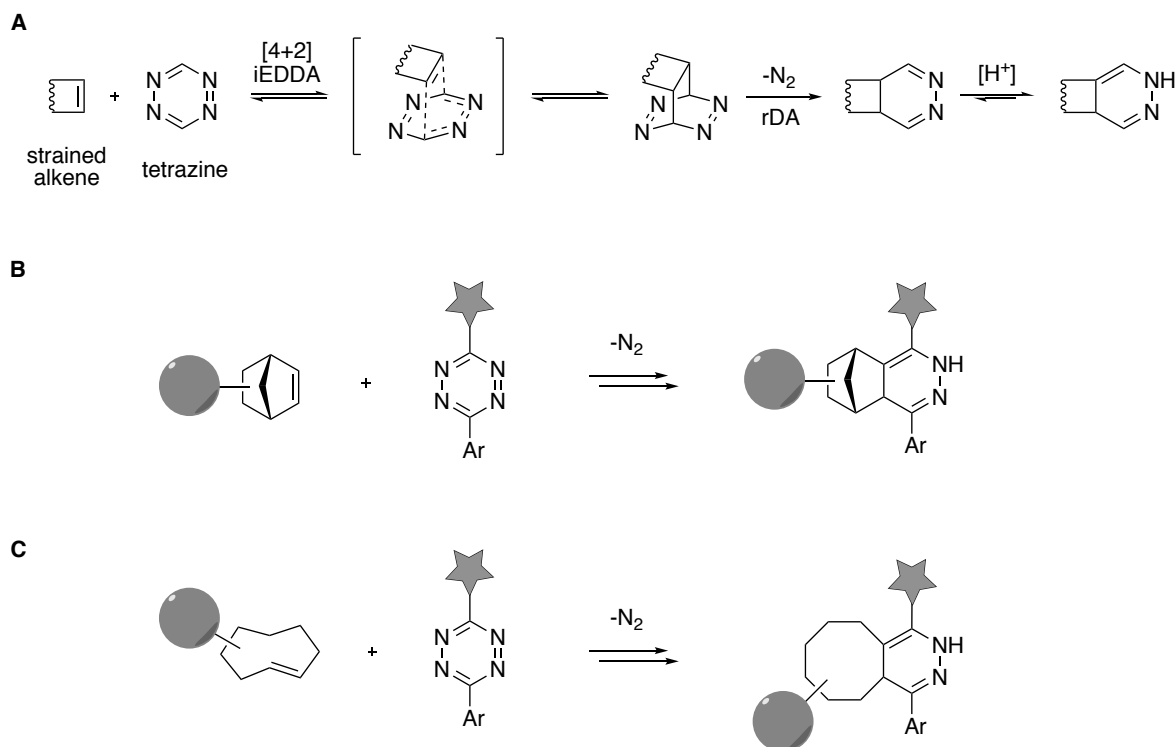


Figure 4.16 Inverse-electron-demand [4+2] Diels-Alder cycloadditions of tetrazines and strained alkenes. **A:** General mechanism. **B:** Norbornene as dienophile. **C:** Trans-cyclooctene as dienophile.

Since then, the bioorthogonal modification of norbornenes has been successfully applied in a number of studies, including site-specific protein modifications by the Carell group.^{1,159,284} Norbornenes are synthetically accessible and their reaction with tetrazines was shown to be chemoselective within living cells and fast enough for *in vivo* experiments (k_2 typically around $2 \text{ M}^{-1}\text{s}^{-1}$). Faster reaction kinetics in inverse electron-demand reactions with second order rates of up to $10^4 \text{ M}^{-1}\text{s}^{-1}$ can be only achieved by trans-cyclooctenes, which have been site specifically introduced into proteins using the pyrrolysine incorporation system as well.^{176,177,280,285-287}

In recent years, mainly because of the unprecedented reaction kinetics and no need for additional catalyst, the inverse electron-demand Diels-Alder cycloaddition became the most promising bioorthogonal labeling reaction for *in vivo* labeling and imaging studies.^{175,288-290} Another feature of these tetrazine-based reactions justifies their newly acquired supremacy in imaging studies: When bound to tetrazines, the fluorescence of

certain fluorophores is quenched through energy transfer. After the reaction of these functionalized tetrazines with alkenes or alkynes, the fluorescence is restored resulting into excellent signal-to-noise ratio, which is highly desirable and advantageous in various bioimaging studies.²⁸⁸ Newest developments by Vrabel et al. even make the attachment of fluorophores potentially obsolete since it could be shown, that the resulting 1,4-dihydropyridazines of axially substituted *trans*-cyclooctenes show strong fluorescence themselves.²⁹¹

4.3.6 Desmethylpyrrolysine-aminobenzaldehyde ligations

The genetic introduction of bioorthogonal reaction handles usually requires the addition of the desired amino acid to the expression medium in millimolar concentrations. Unfortunately, only 0.1% of the NCAA is incorporated into the protein of interest, the majority of the synthesized amino acid is disposed after the protein expression. In case of e.g. visualization studies and the low absolute amount of required labeled protein, the needed amount of NCAA is usually not a problem. In case of expression, purification and *in vitro* modification to gain e.g. antibody drug conjugates or PEGylated therapeutic proteins in big scale, synthetic effort and waste of resources have to be avoided wherever possible. Therefore, Geierstanger et al. utilized the reactivity of desmethylpyrrolysine (dmPyl) which can be biosynthesized by e.g. *E. coli* or mammalian cells when adding D-ornithine to the expression medium and transfected with the needed genes (section 4.1.3.2).²⁹²⁻²⁹⁴ Ornithine can be added as a racemic mixture and is a cheap resource without the need of chemical synthesis.

Desmethylpyrrolysine can be modified chemoselectively using 2-amino-benzaldehydes (2-ABAs) or 2-amino-acetophenones (2-AAPs), a reaction that has been described before for the Pyl-analog pyrroline-5-carboxylate.²⁹² The first step of the conjugation reaction yields a pyrrolo-quinazoline ring which was shown to be stabilized through cyclization of the 2-ABA/2-AAP carbonyl moiety with the ϵ -amide group of dmPyl. This reaction is reversible. The linkage can be fixed by addition of reducing agents such as cyanoborohydride, leading to a stable addition product (Figure 4.17).

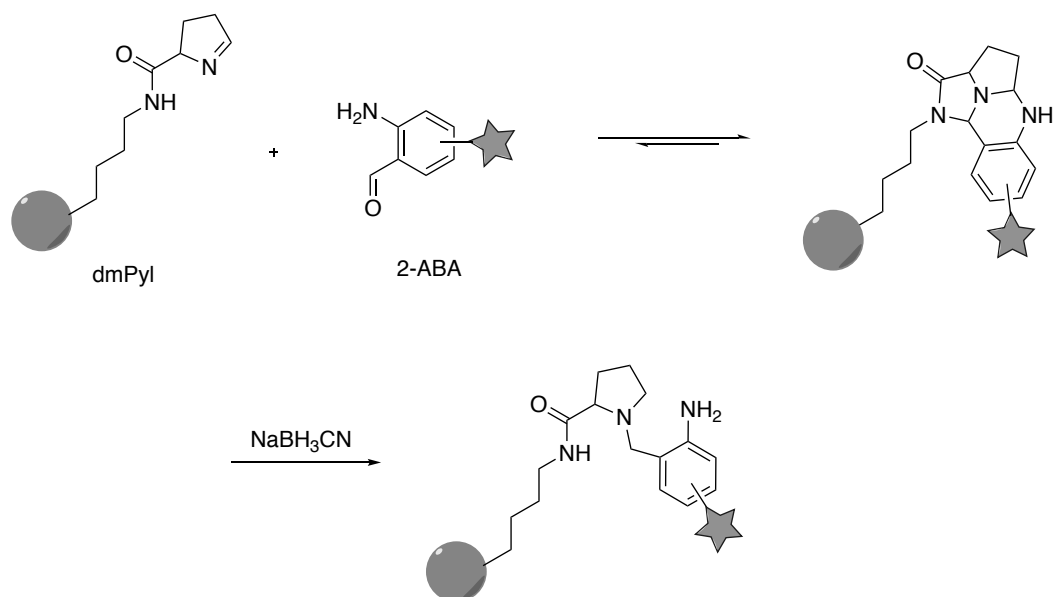


Figure 4.17 Modification of *dmPyl* containing proteins with 2-aminobenzaldehydes (2-ABAs)

4.3.7 Other ligation methods

Other bioorthogonal reactions that have been applied on Pyl analogs include site-specific native chemical ligations, photoactivated click- and crosslinking-reactions or Sonogashira couplings.^{120,181,295,296}

Since the development of bioorthogonal reactions is an area of extensive research, one could expect that new ligation reactions will appear and will be applied on NCAAs introduced by the Pyl incorporation system. Since the requirements of the coupling reaction are not universal but strongly depend on the purpose of the protein conjugate, scientists will benefit from every extension of the bioorthogonal chemistry toolbox.

5 Literature of the introduction

1. Kaya, E., Vrabel, M., Deiml, C., Prill, S., Fluxa, V. S. & Carell, T. A genetically encoded norbornene amino acid for the mild and selective modification of proteins in a copper-free click reaction. *Angew. Chem. Int. Ed.* **51**, 4466–4469 (2012).
2. Agard, N. J., Prescher, J. A. & Bertozzi, C. R. A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems. *J. Am. Chem. Soc.* **126**, 15046–15047 (2004).
3. Baskin, J. M., Prescher, J. A., Laughlin, S. T., Agard, N. J., Chang, P. V., Miller, I. A., Lo, A., Codelli, J. A. & Bertozzi, C. R. Copper-free click chemistry for dynamic in vivo imaging. *Proc. Natl. Acad. Sci.* **104**, 16793–16797 (2007).
4. Lin, X., Yu, A. C. S. & Chan, T. F. Efforts and Challenges in Engineering the Genetic Code. *Life (Basel)* **7**, 12 (2017).
5. Szathmáry, E. The origin of the genetic code: amino acids as cofactors in an RNA world. *Trends Genet.* **15**, 223–229 (1999).
6. Taverna, D. M. & Goldstein, R. A. Why are proteins so robust to site mutations? *J. Mol. Biol.* **315**, 479–484 (2002).
7. Xu, Y. O., Hall, R. W., Goldstein, R. A. & Pollock, D. D. Divergence, recombination and retention of functionality during protein evolution. *Hum. Genomics* **2**, 158–167 (2005).
8. Dufton, M. J. Genetic code synonym quotas and amino acid complexity: cutting the cost of proteins? *J. Theor. Biol.* **187**, 165–173 (1997).
9. Akashi, H. & Gojobori, T. Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*. *Proc. Natl. Acad. Sci.* **99**, 3695–3700 (2002).
10. Walter, K. U., Vamvaca, K. & Hilvert, D. An active enzyme constructed from a 9-amino acid alphabet. *J. Biol. Chem.* **280**, 37742–37746 (2005).
11. Fan, K. & Wang, W. What is the minimum number of letters required to fold a protein? *J. Mol. Biol.* **328**, 921–926 (2003).
12. Crick, F. H. The origin of the genetic code. *J. Mol. Biol.* **38**, 367–379 (1968).
13. Cone, J. E., Del Río, R. M., Davis, J. N. & Stadtman, T. C. Chemical characterization of the selenoprotein component of clostridial glycine reductase: identification of selenocysteine as the organoselenium moiety. *Proc. Natl. Acad. Sci.* **73**, 2659–2663 (1976).
14. Hao, B., Gong, W., Ferguson, T. K., James, C. M., Krzycki, J. A. & Chan, M. K. A new UAG-encoded residue in the structure of a methanogen methyltransferase. *Science* **296**, 1462–1466 (2002).
15. Srinivasan, G., James, C. M. & Krzycki, J. A. Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. *Science* **296**, 1459–1462 (2002).
16. Adachi, M. & Cavalcanti, A. R. O. Tandem stop codons in ciliates that reassign stop codons. *J. Mol. Evol.* **68**, 424–431 (2009).
17. Söll, D. & RajBhandary, U. L. The genetic code - thawing the 'frozen accident'. *J. Biosci.* **31**, 459–463 (2006).
18. Romagné, F., Santesmasses, D., White, L., Sarangi, G. K., Mariotti, M., Hübner, R., Weihmann, A., Parra, G., Gladyshev, V. N., Guigó, R. & Castellano, S. SelenoDB 2.0: annotation of selenoprotein genes in animals and their genetic diversity in humans.

- Nucleic Acids Res.* **42**, D437–43 (2014).
19. Zinoni, F., Birkmann, A., Stadtman, T. C. & Böck, A. Nucleotide sequence and expression of the selenocysteine-containing polypeptide of formate dehydrogenase (formate-hydrogen-lyase-linked) from *Escherichia coli*. *Proc. Natl. Acad. Sci.* **83**, 4650–4654 (1986).
 20. Leinfelder, W., Zehelein, E., Mandrand-Berthelot, M. A. & Böck, A. Gene for a novel tRNA species that accepts L-serine and cotranslationally inserts selenocysteine. *Nature* **331**, 723–725 (1988).
 21. Söll, D. Genetic code: enter a new amino acid. *Nature* **331**, 662–663 (1988).
 22. Forchhammer, K., Leinfelder, W., Boesmiller, K., Veprek, B. & Böck, A. Selenocysteine synthase from *Escherichia coli*. Nucleotide sequence of the gene (*selA*) and purification of the protein. *J. Biol. Chem.* **266**, 6318–6323 (1991).
 23. Forchhammer, K. & Böck, A. Selenocysteine synthase from *Escherichia coli*. Analysis of the reaction sequence. *J. Biol. Chem.* **266**, 6324–6328 (1991).
 24. Berry, M. J., Banu, L., Chen, Y. Y., Mandel, S. J., Kieffer, J. D., Harney, J. W. & Larsen, P. R. Recognition of UGA as a selenocysteine codon in type I deiodinase requires sequences in the 3' untranslated region. *Nature* **353**, 273–276 (1991).
 25. Forchhammer, K., Leinfelder, W. & Böck, A. Identification of a novel translation factor necessary for the incorporation of selenocysteine into protein. *Nature* **342**, 453–456 (1989).
 26. Fagegaltier, D., Hubert, N., Yamada, K., Mizutani, T., Carbon, P. & Krol, A. Characterization of mSelB, a novel mammalian elongation factor for selenoprotein translation. *EMBO J.* **19**, 4796–4805 (2000).
 27. Leinfelder, W., Forchhammer, K., Veprek, B., Zehelein, E. & Böck, A. In vitro synthesis of selenocysteinyl-tRNA(UCA) from seryl-tRNA(UCA): involvement and characterization of the *selD* gene product. *Proc. Natl. Acad. Sci.* **87**, 543–547 (1990).
 28. Allmang, C. & Krol, A. Selenoprotein synthesis: UGA does not end the story. *Biochimie* **88**, 1561–1571 (2006).
 29. Hatfield, D. L., Carlson, B. A., Xu, X.-M., Mix, H. & Gladyshev, V. N. Selenocysteine incorporation machinery and the role of selenoproteins in development and health. *Prog. Nucleic Acid Res. Mol. Biol.* **81**, 97–142 (2006).
 30. Hao, B., Zhao, G., Kang, P. T., Soares, J. A., Ferguson, T. K., Gallucci, J., Krzycki, J. A. & Chan, M. K. Reactivity and chemical synthesis of L-pyrrolysine- the 22(nd) genetically encoded amino acid. *Chem. Biol.* **11**, 1317–1324 (2004).
 31. Soares, J. A., Zhang, L., Pitsch, R. L., Kleinholz, N. M., Jones, R. B., Wolff, J. J., Amster, J., Green-Church, K. B. & Krzycki, J. A. The residue mass of L-pyrrolysine in three distinct methylamine methyltransferases. *J. Biol. Chem.* **280**, 36962–36969 (2005).
 32. Ibba, M. & Söll, D. Genetic code: introducing pyrrolysine. *Curr. Biol.* **12**, R464–6 (2002).
 33. Zhang, Y. & Gladyshev, V. N. High content of proteins containing 21st and 22nd amino acids, selenocysteine and pyrrolysine, in a symbiotic deltaproteobacterium of gutless worm *Olavius algarvensis*. *Nucleic Acids Res.* **35**, 4952–4963 (2007).
 34. Prat, L., Heinemann, I. U., Aerni, H. R., Rinehart, J., O'Donoghue, P. & Söll, D. Carbon source-dependent expansion of the genetic code in bacteria. *Proc. Natl. Acad. Sci.* **109**, 21070–21075 (2012).
 35. Paul, L., Ferguson, D. J. & Krzycki, J. A. The trimethylamine methyltransferase gene and multiple dimethylamine methyltransferase genes of *Methanosarcina barkeri* contain in-

- frame and read-through amber codons. *J. Bacteriol.* **182**, 2520–2529 (2000).
36. Ferguson, D. J., Gorlatova, N., Grahame, D. A. & Krzycki, J. A. Reconstitution of dimethylamine:coenzyme M methyl transfer with a discrete corrinoid protein and two methyltransferases purified from *Methanosarcina barkeri*. *J. Biol. Chem.* **275**, 29053–29060 (2000).
 37. Ferguson, D. J. & Krzycki, J. A. Reconstitution of trimethylamine-dependent coenzyme M methylation with the trimethylamine corrinoid protein and the isozymes of methyltransferase II from *Methanosarcina barkeri*. *J. Bacteriol.* **179**, 846–852 (1997).
 38. Burke, S. A. & Krzycki, J. A. Reconstitution of Monomethylamine:Coenzyme M methyl transfer with a corrinoid protein and two methyltransferases purified from *Methanosarcina barkeri*. *J. Biol. Chem.* **272**, 16570–16577 (1997).
 39. Gaston, M. A., Jiang, R. & Krzycki, J. A. Functional context, biosynthesis, and genetic encoding of pyrrolysine. *Curr. Opin. Microbiol.* **14**, 342–349 (2011).
 40. Longstaff, D. G., Larue, R. C., Faust, J. E., Mahapatra, A., Zhang, L., Green-Church, K. B. & Krzycki, J. A. A natural genetic code expansion cassette enables transmissible biosynthesis and genetic encoding of pyrrolysine. *Proc. Natl. Acad. Sci.* **104**, 1021–1026 (2007).
 41. Hertweck, C. Biosynthesis and charging of pyrrolysine, the 22nd genetically encoded amino acid. *Angew. Chem. Int. Ed.* **50**, 9540–9541 (2011).
 42. Namy, O., Zhou, Y., Gundllapalli, S., Polycarpo, C. R., Denise, A., Rousset, J.-P., Söll, D. & Ambrogelly, A. Adding pyrrolysine to the *Escherichia coli* genetic code. *FEBS Lett.* **581**, 5282–5288 (2007).
 43. Gaston, M. A., Zhang, L., Green-Church, K. B. & Krzycki, J. A. The complete biosynthesis of the genetically encoded amino acid pyrrolysine from lysine. *Nature* **471**, 647–650 (2011).
 44. Cellitti, S. E., Ou, W., Chiu, H.-P., Grünwald, J., Jones, D. H., Hao, X., Fan, Q., Quinn, L. L., Ng, K., Anfora, A. T., Lesley, S. A., Uno, T., Brock, A. & Geierstanger, B. H. D-Ornithine coopts pyrrolysine biosynthesis to make and insert pyrroline-carboxy-lysine. *Nat. Chem. Biol.* **7**, 528–530 (2011).
 45. Quitterer, F., List, A., Eisenreich, W., Bacher, A. & Groll, M. Crystal structure of methylornithine synthase (PylB): insights into the pyrrolysine biosynthesis. *Angew. Chem. Int. Ed.* **51**, 1339–1342 (2012).
 46. Quitterer, F., List, A., Beck, P., Bacher, A. & Groll, M. Biosynthesis of the 22nd genetically encoded amino acid pyrrolysine: structure and reaction mechanism of PylC at 1.5 Å resolution. *J. Mol. Biol.* **424**, 270–282 (2012).
 47. Quitterer, F., Beck, P., Bacher, A. & Groll, M. Structure and reaction mechanism of pyrrolysine synthase (PylD). *Angew. Chem. Int. Ed.* **52**, 7033–7037 (2013).
 48. Fletcher, J. E., Copeland, P. R., Driscoll, D. M. & Krol, A. The selenocysteine incorporation machinery: interactions between the SECIS RNA and the SECIS-binding protein SBP2. *RNA* **7**, 1442–1453 (2001).
 49. Liu, Z., Reches, M., Groisman, I. & Engelberg-Kulka, H. The nature of the minimal ‘selenocysteine insertion sequence’ (SECIS) in *Escherichia coli*. *Nucleic Acids Res.* **26**, 896–902 (1998).
 50. Zhang, Y., Baranov, P. V., Atkins, J. F. & Gladyshev, V. N. Pyrrolysine and selenocysteine use dissimilar decoding strategies. *J. Biol. Chem.* **280**, 20740–20751 (2005).
 51. Ambrogelly, A., Gundllapalli, S., Herring, S., Polycarpo, C., Frauer, C. & Söll, D.

- Pyrrolysine is not hardwired for cotranslational insertion at UAG codons. *Proc. Natl. Acad. Sci.* **104**, 3141–3146 (2007).
52. Young, D. D. & Schultz, P. G. Playing with the Molecules of Life. *ACS Chem. Biol.* **13**, 854–870 (2018).
53. Nguyen, D. P., Garcia Alai, M. M., Kapadnis, P. B., Neumann, H. & Chin, J. W. Genetically encoding N(epsilon)-methyl-L-lysine in recombinant histones. *J. Am. Chem. Soc.* **131**, 14194–14195 (2009).
54. Neumann, H., Peak-Chew, S. Y. & Chin, J. W. Genetically encoding N(epsilon)-acetyllysine in recombinant proteins. *Nat. Chem. Biol.* **4**, 232–234 (2008).
55. Neumann, H., Hancock, S. M., Buning, R., Routh, A., Chapman, L., Somers, J., Owen-Hughes, T., van Noort, J., Rhodes, D. & Chin, J. W. A method for genetically installing site-specific acetylation in recombinant histones defines the effects of H3 K56 acetylation. *Mol. Cell* **36**, 153–163 (2009).
56. Wang, Z. A., Kurra, Y., Wang, X., Zeng, Y., Lee, Y.-J., Sharma, V., Lin, H., Dai, S. Y. & Liu, W. R. A Versatile Approach for Site-Specific Lysine Acylation in Proteins. *Angew. Chem. Int. Ed.* **56**, 1643–1647 (2017).
57. Wang, W. W., Zeng, Y., Wu, B., Deiters, A. & Liu, W. R. A Chemical Biology Approach to Reveal Sirt6-targeted Histone H3 Sites in Nucleosomes. *ACS Chem. Biol.* **11**, 1973–1981 (2016).
58. Xuan, W., Yao, A. & Schultz, P. G. Genetically Encoded Fluorescent Probe for Detecting Sirtuins in Living Cells. *J. Am. Chem. Soc.* **139**, 12350–12353 (2017).
59. Luo, X., Fu, G., Wang, R. E., Zhu, X., Zambaldo, C., Liu, R., Liu, T., Lyu, X., Du, J., Xuan, W., Yao, A., Reed, S. A., Kang, M., Zhang, Y., Guo, H., Huang, C., Yang, P.-Y., Wilson, I. A., Schultz, P. G. & Wang, F. Genetically encoding phosphotyrosine and its nonhydrolyzable analog in bacteria. *Nat. Chem. Biol.* **13**, 845–849 (2017).
60. Liu, C. C. & Schultz, P. G. Adding new chemistries to the genetic code. *Annu. Rev. Biochem.* **79**, 413–444 (2010).
61. Li, J. C., Liu, T., Wang, Y., Mehta, A. P. & Schultz, P. G. Enhancing Protein Stability with Genetically Encoded Noncanonical Amino Acids. *J. Am. Chem. Soc.* **140**, 15997–16000 (2018).
62. Liu, T., Wang, Y., Luo, X., Li, J., Reed, S. A., Xiao, H., Young, T. S. & Schultz, P. G. Enhancing protein stability with extended disulfide bonds. *Proc. Natl. Acad. Sci.* **113**, 5910–5915 (2016).
63. Oyala, P. H., Ravichandran, K. R., Funk, M. A., Stucky, P. A., Stich, T. A., Drennan, C. L., Britt, R. D. & Stubbe, J. Biophysical Characterization of Fluorotyrosine Probes Site-Specifically Incorporated into Enzymes: E. coli Ribonucleotide Reductase As an Example. *J. Am. Chem. Soc.* **138**, 7951–7964 (2016).
64. Villa, J. K., Tran, H.-A., Vipani, M., Gianturco, S., Bhasin, K., Russell, B. L., Harbron, E. J. & Young, D. D. Fluorescence Modulation of Green Fluorescent Protein Using Fluorinated Unnatural Amino Acids. *Molecules* **22**, 1194 (2017).
65. Liu, J., Hemphill, J., Samanta, S., Tsang, M. & Deiters, A. Genetic Code Expansion in Zebrafish Embryos and Its Application to Optical Control of Cell Signaling. *J. Am. Chem. Soc.* **139**, 9100–9103 (2017).
66. Koh, M., Nasertorabi, F., Han, G. W., Stevens, R. C. & Schultz, P. G. Generation of an Orthogonal Protein-Protein Interface with a Noncanonical Amino Acid. *J. Am. Chem. Soc.* **jacs.7b02273** (2017). doi:10.1021/jacs.7b02273
67. Rannversson, H., Andersen, J., Sørensen, L., Bang-Andersen, B., Park, M., Huber, T.,

- Sakmar, T. P. & Strømgaard, K. Genetically encoded photocrosslinkers locate the high-affinity binding site of antidepressant drugs in the human serotonin transporter. *Nat Commun* **7**, 11261 (2016).
68. Xuan, W., Shao, S. & Schultz, P. G. Protein Crosslinking by Genetically Encoded Noncanonical Amino Acids with Reactive Aryl Carbamate Side Chains. *Angew. Chem. Int. Ed.* **56**, 5096–5100 (2017).
69. Mitchell, A. L., Addy, P. S., Chin, M. A. & Chatterjee, A. A Unique Genetically Encoded FRET Pair in Mammalian Cells. *ChemBioChem* **18**, 511–514 (2017).
70. Lv, X., Yu, Y., Zhou, M., Hu, C., Gao, F., Li, J., Liu, X., Deng, K., Zheng, P., Gong, W., Xia, A. & Wang, J. Ultrafast photoinduced electron transfer in green fluorescent protein bearing a genetically encoded electron acceptor. *J. Am. Chem. Soc.* **137**, 7270–7273 (2015).
71. Basom, E. J., Maj, M., Cho, M. & Thielges, M. C. Site-Specific Characterization of Cytochrome P450cam Conformations by Infrared Spectroscopy. *Anal. Chem.* **88**, 6598–6606 (2016).
72. Mills, J. H., Sheffler, W., Ener, M. E., Almhjell, P. J., Oberdorfer, G., Pereira, J. H., Parmeggiani, F., Sankaran, B., Zwart, P. H. & Baker, D. Computational design of a homotrimeric metalloprotein with a trisbipyridyl core. *Proc. Natl. Acad. Sci.* **113**, 15012–15017 (2016).
73. Luo, J., Liu, Q., Morihiro, K. & Deiters, A. Small-molecule control of protein function through Staudinger reduction. *Nat Chem* **8**, 1027–1034 (2016).
74. Basak, S., Punetha, V. D., Bisht, G., Bisht, S. S., Sahoo, N. G. & Cho, J. W. Recent Trends of Polymer-Protein Conjugate Application in Biocatalysis: A Review. *Polymer Reviews* **55**, 163–198 (2015).
75. Baslé, E., Joubert, N. & Pucheault, M. Protein chemical modification on endogenous amino acids. *Chem. Biol.* **17**, 213–227 (2010).
76. Kim Wals, H. O. Unnatural amino acid incorporation in *E. coli*: current and future applications in the design of therapeutic proteins. *Frontiers in Chemistry* **2**, 237 (2014).
77. Peng, T. & Hang, H. C. Site-Specific Bioorthogonal Labeling for Fluorescence Imaging of Intracellular Proteins in Living Cells. *J. Am. Chem. Soc.* **138**, 14423–14433 (2016).
78. Pisal, D. S., Kosloski, M. P. & Balu-Iyer, S. V. Delivery of therapeutic proteins. *J Pharm Sci* **99**, 2557–2575 (2010).
79. Payne, R. W., Murphy, B. M. & Manning, M. C. Product development issues for PEGylated proteins. *Pharm Dev Technol* **16**, 423–440 (2011).
80. Biganzoli, L., Untch, M., Skacel, T. & Pico, J.-L. Neulasta (pegfilgrastim): a once-per-cycle option for the management of chemotherapy-induced neutropenia. *Semin. Oncol.* **31**, 27–34 (2004).
81. Molineux, G. The design and development of pegfilgrastim (PEG-rmetHuG-CSF, Neulasta). *Curr. Pharm. Des.* **10**, 1235–1244 (2004).
82. Senter, P. D. Potent antibody drug conjugates for cancer therapy. *Curr Opin Chem Biol* **13**, 235–244 (2009).
83. Chari, R. V. J. Targeted cancer therapy: conferring specificity to cytotoxic drugs. *Acc. Chem. Res.* **41**, 98–107 (2008).
84. Parslow, A. C., Parakh, S., Lee, F.-T., Gan, H. K. & Scott, A. M. Antibody-Drug Conjugates for Cancer Therapy. *Biomedicines* **4**, 14 (2016).

85. VanBrunt, M. P., Shanebeck, K., Caldwell, Z., Johnson, J., Thompson, P., Martin, T., Dong, H., Li, G., Xu, H., D'Hooge, F., Masterson, L., Bariola, P., Tiberghien, A., Ezeadi, E., Williams, D. G., Hartley, J. A., Howard, P. W., Grabstein, K. H., Bowen, M. A. & Marelli, M. Genetically Encoded Azide Containing Amino Acid in Mammalian Cells Enables Site-Specific Antibody-Drug Conjugates Using Click Cycloaddition Chemistry. *Bioconjug. Chem.* **26**, 2249–2260 (2015).
86. LoRusso, P. M., Weiss, D., Guardino, E., Girish, S. & Sliwkowski, M. X. Trastuzumab emtansine: a unique antibody-drug conjugate in development for human epidermal growth factor receptor 2-positive cancer. *Clin. Cancer Res.* **17**, 6437–6447 (2011).
87. Teicher, B. A. & Doroshow, J. H. The promise of antibody-drug conjugates. *N. Engl. J. Med.* **367**, 1847–1848 (2012).
88. Verma, S., Miles, D., Gianni, L., Krop, I. E., Welslau, M., Baselga, J., Pegram, M., Oh, D.-Y., Diéras, V., Guardino, E., Fang, L., Lu, M. W., Olsen, S., Blackwell, K. EMILIA Study Group. Trastuzumab emtansine for HER2-positive advanced breast cancer. *N. Engl. J. Med.* **367**, 1783–1791 (2012).
89. Hu, Q.-Y., Berti, F. & Adamo, R. Towards the next generation of biomedicines by site-selective conjugation. *Chem Soc Rev* **45**, 1691–1719 (2016).
90. Turecek, P. L., Bossard, M. J., Schoetens, F. & Ivens, I. A. PEGylation of Biopharmaceuticals: A Review of Chemistry and Nonclinical Safety Information of Approved Drugs. *J Pharm Sci* **105**, 460–475 (2016).
91. Griffin, B. A., Adams, S. R. & Tsien, R. Y. Specific covalent labeling of recombinant protein molecules inside live cells. *Science* **281**, 269–272 (1998).
92. Lai, Y.-T., Chang, Y.-Y., Hu, L., Yang, Y., Chao, A., Du, Z.-Y., Tanner, J. A., Chye, M.-L., Qian, C., Ng, K.-M., Li, H. & Sun, H. Rapid labeling of intracellular His-tagged proteins in living cells. *Proc. Natl. Acad. Sci.* **112**, 2948–2953 (2015).
93. Keppler, A., Gendreizig, S., Gronemeyer, T., Pick, H., Vogel, H. & Johnsson, K. A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nat. Biotechnol.* **21**, 86–89 (2003).
94. Gautier, A., Juillerat, A., Heinis, C., Corrêa, I. R., Kindermann, M., Beaufils, F. & Johnsson, K. An engineered protein tag for multiprotein labeling in living cells. *Chem. Biol.* **15**, 128–136 (2008).
95. Carrico, I. S., Carlson, B. L. & Bertozzi, C. R. Introducing genetically encoded aldehydes into proteins. *Nat. Chem. Biol.* **3**, 321–322 (2007).
96. Merrifield, R. B. *Solid-phase peptide synthesis*. *Adv. Enzymol Relat. Areas Mci.* (Biol, 1969).
97. James S Nowick, Kit S Lam, Tatyana V Khasanova, William E Kemnitzer, Santanu Maitra, Hao T Mee, A. & Liu, R. *An Unnatural Amino Acid that Induces β -Sheet Folding and Interaction in Peptides*. *J. Am. Chem. Soc.* **124**, 4972–4973 (American Chemical Society, 2002).
98. Dawson, P. E. & Kent, S. B. Synthesis of native proteins by chemical ligation. *Annu. Rev. Biochem.* **69**, 923–960 (2000).
99. Muir, T. W. Semisynthesis of proteins by expressed protein ligation. *Annu. Rev. Biochem.* **72**, 249–289 (2003).
100. Nilsson, B. L., Soellner, M. B. & Raines, R. T. Chemical synthesis of proteins. *Annu Rev Biophys Biomol Struct* **34**, 91–118 (2005).
101. David, R., Richter, M. P. O. & Beck-Sickinger, A. G. Expressed protein ligation. Method

- and applications. *Eur. J. Biochem.* **271**, 663–677 (2004).
102. Bain, J. D., Diala, E. S., Glabe, C. G., Wacker, D. A., Lyttle, M. H., Dix, T. A. & Chamberlin, A. R. Site-specific incorporation of nonnatural residues during in vitro protein biosynthesis with semisynthetic aminoacyl-tRNAs. *Biochemistry* **30**, 5411–5421 (1991).
103. Leisle, L., Valiyaveetil, F., Mehl, R. A. & Ahern, C. A. Incorporation of Non-Canonical Amino Acids. *Adv. Exp. Med. Biol.* **869**, 119–151 (2015).
104. Bohlke, N. & Budisa, N. Sense codon emancipation for proteome-wide incorporation of noncanonical amino acids: rare isoleucine codon AUA as a target for genetic code expansion. *FEMS Microbiol. Lett.* **351**, 133–144 (2014).
105. Mukai, T., Yamaguchi, A., Ohtake, K., Takahashi, M., Hayashi, A., Iraha, F., Kira, S., Yanagisawa, T., Yokoyama, S., Hoshi, H., Kobayashi, T. & Sakamoto, K. Reassignment of a rare sense codon to a non-canonical amino acid in *Escherichia coli*. *Nucleic Acids Res.* **43**, 8111–8122 (2015).
106. Zeng, Y., Wang, W. & Liu, W. R. Towards reassigning the rare AGG codon in *Escherichia coli*. *ChemBioChem* **15**, 1750–1754 (2014).
107. Wang, Y. & Tsao, M.-L. Reassigning Sense Codon AGA to Encode Noncanonical Amino Acids in *Escherichia coli*. *ChemBioChem* **17**, 2234–2239 (2016).
108. Wang, K., Schmied, W. H. & Chin, J. W. Reprogramming the genetic code: from triplet to quadruplet codes. *Angew. Chem. Int. Ed.* **51**, 2288–2297 (2012).
109. Anderson, J. C., Wu, N., Santoro, S. W., Lakshman, V., King, D. S. & Schultz, P. G. An expanded genetic code with a functional quadruplet codon. *Proc. Natl. Acad. Sci.* **101**, 7566–7571 (2004).
110. Rodriguez, E. A., Lester, H. A. & Dougherty, D. A. In vivo incorporation of multiple unnatural amino acids through nonsense and frameshift suppression. *Proc. Natl. Acad. Sci.* **103**, 8650–8655 (2006).
111. Niu, W., Schultz, P. G. & Guo, J. An expanded genetic code in mammalian cells with a functional quadruplet codon. *ACS Chem. Biol.* **8**, 1640–1645 (2013).
112. Magliery, T. J., Anderson, J. C. & Schultz, P. G. Expanding the genetic code: selection of efficient suppressors of four-base codons and identification of ‘shifty’ four-base codons with a library approach in *Escherichia coli*. *J. Mol. Biol.* **307**, 755–769 (2001).
113. Anderson, J. C., Magliery, T. J. & Schultz, P. G. Exploring the limits of codon and anticodon size. *Chem. Biol.* **9**, 237–244 (2002).
114. Taki, M., Matsushita, J. & Sisido, M. Expanding the genetic code in a mammalian cell line by the introduction of four-base codon/anticodon pairs. *ChemBioChem* **7**, 425–428 (2006).
115. Nakamura, Y., Gojobori, T. & Ikemura, T. Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res.* **28**, 292 (2000).
116. Xie, J. & Schultz, P. G. An expanding genetic code. *Methods* **36**, 227–238 (2005).
117. Mukai, T., Hoshi, H., Ohtake, K., Takahashi, M., Yamaguchi, A., Hayashi, A., Yokoyama, S. & Sakamoto, K. Highly reproductive *Escherichia coli* cells with no specific assignment to the UAG codon. *Sci Rep* **5**, 9699 (2015).
118. Wang, L. & Schultz, P. G. A general approach for the generation of orthogonal tRNAs. *Chem. Biol.* **8**, 883–890 (2001).
119. Italia, J. S., Addy, P. S., Wrobel, C. J. J., Crawford, L. A., Lajoie, M. J., Zheng, Y. & Chatterjee, A. An orthogonalized platform for genetic code expansion in both bacteria and

- eukaryotes. *Nat. Chem. Biol.* **13**, 446–450 (2017).
120. Wang, L., Brock, A., Herberich, B. & Schultz, P. G. Expanding the genetic code of *Escherichia coli*. *Science* **292**, 498–500 (2001).
121. Chin, J. W. Expanding and reprogramming the genetic code of cells and animals. *Annu. Rev. Biochem.* **83**, 379–408 (2014).
122. Wang, L., Zhang, Z., Brock, A. & Schultz, P. G. Addition of the keto functional group to the genetic code of *Escherichia coli*. *Proc. Natl. Acad. Sci.* **100**, 56–61 (2003).
123. Zhang, Z., Smith, B. A. C., Wang, L., Brock, A., Cho, C. & Schultz, P. G. A new strategy for the site-specific modification of proteins in vivo. *Biochemistry* **42**, 6735–6746 (2003).
124. Chin, J. W., Cropp, T. A., Anderson, J. C., Mukherji, M., Zhang, Z. & Schultz, P. G. An expanded eukaryotic genetic code. *Science* **301**, 964–967 (2003).
125. Young, T. S., Ahmad, I., Brock, A. & Schultz, P. G. Expanding the genetic repertoire of the methylotrophic yeast *Pichia pastoris*. *Biochemistry* **48**, 2643–2653 (2009).
126. Liu, W., Brock, A., Chen, S., Chen, S. & Schultz, P. G. Genetic incorporation of unnatural amino acids into proteins in mammalian cells. *Nat. Methods* **4**, 239–244 (2007).
127. Zeng, H., Xie, J. & Schultz, P. G. Genetic introduction of a diketone-containing amino acid into proteins. *Bioorg. Med. Chem. Lett.* **16**, 5356–5359 (2006).
128. Zhang, Z., Wang, L., Brock, A. & Schultz, P. G. The selective incorporation of alkenes into proteins in *Escherichia coli*. *Angew. Chem. Int. Ed.* **41**, 2840–2842 (2002).
129. Deiters, A. & Schultz, P. G. In vivo incorporation of an alkyne into proteins in *Escherichia coli*. *Bioorg. Med. Chem. Lett.* **15**, 1521–1524 (2005).
130. Chin, J. W., Santoro, S. W., Martin, A. B., King, D. S., Wang, L. & Schultz, P. G. Addition of p-azido-L-phenylalanine to the genetic code of *Escherichia coli*. *J. Am. Chem. Soc.* **124**, 9026–9027 (2002).
131. Deiters, A., Cropp, T. A., Mukherji, M., Chin, J. W., Anderson, J. C. & Schultz, P. G. Adding amino acids with novel reactivity to the genetic code of *Saccharomyces cerevisiae*. *J. Am. Chem. Soc.* **125**, 11782–11783 (2003).
132. Alfonta, L., Zhang, Z., Uryu, S., Loo, J. A. & Schultz, P. G. Site-specific incorporation of a redox-active amino acid into proteins. *J. Am. Chem. Soc.* **125**, 14662–14663 (2003).
133. Tippmann, E. M. & Schultz, P. G. A genetically encoded metallocene containing amino acid. *Tetrahedron* **63**, 6182–6184 (2007).
134. Xie, J., Liu, W. & Schultz, P. G. A genetically encoded bidentate, metal-binding amino acid. *Angew. Chem. Int. Ed.* **46**, 9239–9242 (2007).
135. Chin, J. W., Martin, A. B., King, D. S., Wang, L. & Schultz, P. G. Addition of a photocrosslinking amino acid to the genetic code of *Escherichia coli*. *Proc. Natl. Acad. Sci.* **99**, 11020–11024 (2002).
136. Wang, W., Takimoto, J. K., Louie, G. V., Baiga, T. J., Noel, J. P., Lee, K.-F., Slesinger, P. A. & Wang, L. Genetically encoding unnatural amino acids for cellular and neuronal studies. *Nat. Neurosci.* **10**, 1063–1072 (2007).
137. Hino, N., Okazaki, Y., Kobayashi, T., Hayashi, A., Sakamoto, K. & Yokoyama, S. Protein photo-cross-linking in mammalian cells by site-specific incorporation of a photoreactive amino acid. *Nat. Methods* **2**, 201–206 (2005).
138. Bose, M., Groff, D., Xie, J., Brustad, E. & Schultz, P. G. The incorporation of a

- photoisomerizable amino acid into proteins in *E. coli*. *J. Am. Chem. Soc.* **128**, 388–389 (2006).
139. Pastrnak, M., Magliery, T. J. & Schultz, P. G. A new orthogonal suppressor tRNA/aminoacyl-tRNA synthetase pair for evolving an organism with an expanded genetic code. *Helvetica Chimica Acta* (2000).
140. Liu, D. R., Magliery, T. J., Pastrnak, M. & Schultz, P. G. Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins in vivo. *Proc. Natl. Acad. Sci.* **94**, 10092–10097 (1997).
141. Ohno, S., Yokogawa, T., Fujii, I., Asahara, H., Inokuchi, H. & Nishikawa, K. Co-expression of yeast amber suppressor tRNA^{Tyr} and tyrosyl-tRNA synthetase in *Escherichia coli*: possibility to expand the genetic code. *J. Biochem.* **124**, 1065–1068 (1998).
142. Furter, R. Expansion of the genetic code: site-directed p-fluoro-phenylalanine incorporation in *Escherichia coli*. *Protein Sci.* **7**, 419–426 (1998).
143. Kowal, A. K., Kohrer, C. & RajBhandary, U. L. Twenty-first aminoacyl-tRNA synthetase-suppressor tRNA pairs for possible use in site-specific incorporation of amino acid analogues into proteins in eukaryotes and in eubacteria. *Proc. Natl. Acad. Sci.* **98**, 2268–2273 (2001).
144. Anderson, J. C. & Schultz, P. G. Adaptation of an orthogonal archaeal leucyl-tRNA and synthetase pair for four-base, amber, and opal suppression. *Biochemistry* **42**, 9598–9608 (2003).
145. Blight, S. K., Larue, R. C., Mahapatra, A., Longstaff, D. G., Chang, E., Zhao, G., Kang, P. T., Green-Church, K. B., Chan, M. K. & Krzycki, J. A. Direct charging of tRNA(CUA) with pyrrolysine in vitro and in vivo. *Nature* **431**, 333–335 (2004).
146. Bianco, A., Townsley, F. M., Greiss, S., Lang, K. & Chin, J. W. Expanding the genetic code of *Drosophila melanogaster*. *Nat. Chem. Biol.* **8**, 748–750 (2012).
147. Ernst, R. J., Krogager, T. P., Maywood, E. S., Zanchi, R., Beránek, V., Elliott, T. S., Barry, N. P., Hastings, M. H. & Chin, J. W. Genetic code expansion in the mouse brain. *Nat. Chem. Biol.* **12**, 776–778 (2016).
148. Yanagisawa, T., Ishii, R., Fukunaga, R., Kobayashi, T., Sakamoto, K. & Yokoyama, S. Multistep engineering of pyrrolysyl-tRNA synthetase to genetically encode N(epsilon)-(o-azidobenzoyloxycarbonyl) lysine for site-specific protein modification. *Chem. Biol.* **15**, 1187–1197 (2008).
149. Nozawa, K., O'Donoghue, P., Gundllapalli, S., Arais, Y., Ishitani, R., Umehara, T., Söll, D. & Nureki, O. Pyrrolysyl-tRNA synthetase-tRNA(Pyl) structure reveals the molecular basis of orthogonality. *Nature* **457**, 1163–1167 (2009).
150. Mukai, T., Kobayashi, T., Hino, N., Yanagisawa, T., Sakamoto, K. & Yokoyama, S. Adding l-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases. *Biochem. Biophys. Res. Commun.* **371**, 818–822 (2008).
151. Polycarpo, C. R., Herring, S., Bérubé, A., Wood, J. L., Söll, D. & Ambrogelly, A. Pyrrolysine analogues as substrates for pyrrolysyl-tRNA synthetase. *FEBS Lett.* **580**, 6695–6700 (2006).
152. Nguyen, D. P., Garcia Alai, M. M., Virdee, S. & Chin, J. W. Genetically directing ε-N, N-dimethyl-L-lysine in recombinant histones. *Chem. Biol.* **17**, 1072–1076 (2010).
153. Virdee, S., Ye, Y., Nguyen, D. P., Komander, D. & Chin, J. W. Engineered diubiquitin synthesis reveals Lys29-isopeptide specificity of an OTU deubiquitinase. *Nat. Chem. Biol.* **6**, 750–757 (2010).

-
154. Li, W.-T., Mahapatra, A., Longstaff, D. G., Bechtel, J., Zhao, G., Kang, P. T., Chan, M. K. & Krzycki, J. A. Specificity of pyrrolysyl-tRNA synthetase for pyrrolysine and pyrrolysine analogs. *J. Mol. Biol.* **385**, 1156–1164 (2009).
155. Fekner, T., Li, X., Lee, M. M. & Chan, M. K. A pyrrolysine analogue for protein click chemistry. *Angew. Chem. Int. Ed.* **48**, 1633–1635 (2009).
156. Li, X., Fekner, T. & Chan, M. K. N6-(2-(R)-propargylglycyl)lysine as a clickable pyrrolysine mimic. *Chem Asian J* **5**, 1765–1769 (2010).
157. Lee, M. M., Fekner, T., Tang, T.-H., Wang, L., Chan, A. H.-Y., Hsu, P.-H., Au, S. W. & Chan, M. K. A click-and-release pyrrolysine analogue. *ChemBioChem* **14**, 805–808 (2013).
158. Nguyen, D. P., Lusic, H., Neumann, H., Kapadnis, P. B., Deiters, A. & Chin, J. W. Genetic encoding and labeling of aliphatic azides and alkynes in recombinant proteins via a pyrrolysyl-tRNA Synthetase/tRNA(CUA) pair and click chemistry. *J. Am. Chem. Soc.* **131**, 8720–8721 (2009).
159. Lang, K., Davis, L., Torres-Kolbus, J., Chou, C., Deiters, A. & Chin, J. W. Genetically encoded norbornene directs site-specific cellular protein labelling via a rapid bioorthogonal reaction. *Nat Chem* **4**, 298–304 (2012).
160. Li, X., Fekner, T., Ottesen, J. J. & Chan, M. K. A pyrrolysine analogue for site-specific protein ubiquitination. *Angew. Chem. Int. Ed.* **48**, 9184–9187 (2009).
161. Nguyen, D. P., Elliott, T., Holt, M., Muir, T. W. & Chin, J. W. Genetically encoded 1,2-aminothiols facilitate rapid and site-specific protein labeling via a bio-orthogonal cyanobenzothiazole condensation. *J. Am. Chem. Soc.* **133**, 11418–11421 (2011).
162. Ai, H.-W., Lee, J. W. & Schultz, P. G. A method to site-specifically introduce methyllysine into proteins in *E. coli*. *Chem. Commun.* **46**, 5506–5508 (2010).
163. Chou, C., Uprety, R., Davis, L., Chin, J. W. & Deiters, A. Genetically encoding an aliphatic diazirine for protein photocrosslinking. *Chem. Sci.* **2**, 480–483 (2011).
164. Wong, M. L., Guzei, I. A. & Kiessling, L. L. An asymmetric synthesis of L-pyrrolysine. *Org. Lett.* **14**, 1378–1381 (2012).
165. Li, Y., Yang, M., Huang, Y., Song, X., Liu, L. & Chen, P. R. Genetically encoded alkenyl-pyrrolysine analogues for thiol-ene reaction mediated site-specific protein labeling. *Chem. Sci.* **3**, 2766–2770 (2012).
166. Yanagisawa, T., Ishii, R., Fukunaga, R., Kobayashi, T., Sakamoto, K. & Yokoyama, S. Crystallographic studies on multiple conformational states of active-site loops in pyrrolysyl-tRNA synthetase. *J. Mol. Biol.* **378**, 634–652 (2008).
167. Kavran, J. M., Gundllapalli, S., O'Donoghue, P., Englert, M., Söll, D. & Steitz, T. A. Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation. *Proc. Natl. Acad. Sci.* **104**, 11268–11273 (2007).
168. Yanagisawa, T., Sumida, T., Ishii, R. & Yokoyama, S. A novel crystal form of pyrrolysyl-tRNA synthetase reveals the pre- and post-aminoacyl-tRNA synthesis conformational states of the adenylate and aminoacyl moieties and an asparagine residue in the catalytic site. *Acta Crystallogr. D Biol. Crystallogr.* **69**, 5–15 (2013).
169. Delarue, M. Aminoacyl-tRNA synthetases. *Curr. Opin. Struct. Biol.* **5**, 48–55 (1995).
170. Ibba, M., Thomann, H. U., Hong, K. W., Sherman, J. M., Weygand-Durasevic, I., Sever, S., Stange-Thomann, N., Praetorius, M. & Söll, D. Substrate selection by aminoacyl-tRNA synthetases. *Nucleic Acids Symp. Ser.* 40–42 (1995).

-
171. Woese, C. R., Olsen, G. J., Ibba, M. & Söll, D. Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. *Microbiol. Mol. Biol. Rev.* **64**, 202–236 (2000).
172. Bryson, D. I., Fan, C., Guo, L.-T., Miller, C., Söll, D. & Liu, D. R. Continuous directed evolution of aminoacyl-tRNA synthetases. *Nat. Chem. Biol.* **13**, 1253–1260 (2017).
173. Hao, Z., Song, Y., Lin, S., Yang, M., Liang, Y., Wang, J. & Chen, P. R. A readily synthesized cyclic pyrrolysine analogue for site-specific protein ‘click’ labeling. *Chem. Commun.* **47**, 4502–4504 (2011).
174. Plass, T., Milles, S., Koehler, C., Schultz, C. & Lemke, E. A. Genetically encoded copper-free click chemistry. *Angew. Chem. Int. Ed.* **50**, 3878–3881 (2011).
175. Borrmann, A., Milles, S., Plass, T., Dommerholt, J., Verkade, J. M. M., Wiessler, M., Schultz, C., van Hest, J. C. M., van Delft, F. L. & Lemke, E. A. Genetic encoding of a bicyclo[6.1.0]nonyne-charged amino acid enables fast cellular protein imaging by metal-free ligation. *ChemBioChem* **13**, 2094–2099 (2012).
176. Plass, T., Milles, S., Koehler, C., Szymański, J., Mueller, R., Wiessler, M., Schultz, C. & Lemke, E. A. Amino acids for Diels-Alder reactions in living cells. *Angew. Chem. Int. Ed.* **51**, 4166–4170 (2012).
177. Lang, K., Davis, L., Wallace, S., Mahesh, M., Cox, D. J., Blackman, M. L., Fox, J. M. & Chin, J. W. Genetic Encoding of bicyclononynes and trans-cyclooctenes for site-specific protein labeling in vitro and in live mammalian cells via rapid fluorogenic Diels-Alder reactions. *J. Am. Chem. Soc.* **134**, 10317–10320 (2012).
178. Lee, Y.-J., Wu, B., Raymond, J. E., Zeng, Y., Fang, X., Wooley, K. L. & Liu, W. R. A genetically encoded acrylamide functionality. *ACS Chem. Biol.* **8**, 1664–1670 (2013).
179. Li, F., Zhang, H., Sun, Y., Pan, Y., Zhou, J. & Wang, J. Expanding the genetic code for photoclick chemistry in *E. coli*, mammalian cells, and *A. thaliana*. *Angew. Chem. Int. Ed.* **52**, 9700–9704 (2013).
180. Zhang, M., Lin, S., Song, X., Liu, J., Fu, Y., Ge, X., Fu, X., Chang, Z. & Chen, P. R. A genetically incorporated crosslinker reveals chaperone cooperation in acid resistance. *Nat. Chem. Biol.* **7**, 671–677 (2011).
181. Lin, S., Zhang, Z., Xu, H., Li, L., Chen, S., Li, J., Hao, Z. & Chen, P. R. Site-specific incorporation of photo-cross-linker and bioorthogonal amino acids into enteric bacterial pathogens. *J. Am. Chem. Soc.* **133**, 20581–20587 (2011).
182. Ai, H.-W., Shen, W., Sagi, A., Chen, P. R. & Schultz, P. G. Probing protein-protein interactions with a genetically encoded photo-crosslinking amino acid. *ChemBioChem* **12**, 1854–1857 (2011).
183. Gautier, A., Nguyen, D. P., Lusic, H., An, W., Deiters, A. & Chin, J. W. Genetically encoded photocontrol of protein localization in mammalian cells. *J. Am. Chem. Soc.* **132**, 4086–4088 (2010).
184. Gautier, A., Deiters, A. & Chin, J. W. Light-activated kinases enable temporal dissection of signaling networks in living cells. *J. Am. Chem. Soc.* **133**, 2124–2127 (2011).
185. Hemphill, J., Chou, C., Chin, J. W. & Deiters, A. Genetically encoded light-activated transcription for spatiotemporal control of gene expression and gene silencing in mammalian cells. *J. Am. Chem. Soc.* **135**, 13433–13439 (2013).
186. Groff, D., Chen, P. R., Peters, F. B. & Schultz, P. G. A genetically encoded epsilon-N-methyl lysine in mammalian cells. *ChemBioChem* **11**, 1066–1068 (2010).

-
187. Wang, Y.-S., Wu, B., Wang, Z., Huang, Y., Wan, W., Russell, W. K., Pai, P.-J., Moe, Y. N., Russell, D. H. & Liu, W. R. A genetically encoded photocaged Nepsilon-methyl-L-lysine. *Mol Biosyst* **6**, 1557–1560 (2010).
188. Huang, Y., Russell, W. K., Wan, W., Pai, P.-J., Russell, D. H. & Liu, W. A convenient method for genetic incorporation of multiple noncanonical amino acids into one protein in *Escherichia coli*. *Mol Biosyst* **6**, 683–686 (2010).
189. Liu, W. R., Wang, Y.-S. & Wan, W. Synthesis of proteins with defined posttranslational modifications using the genetic noncanonical amino acid incorporation approach. *Mol Biosyst* **7**, 38–47 (2011).
190. Kim, C. H., Kang, M., Kim, H. J., Chatterjee, A. & Schultz, P. G. Site-specific incorporation of ϵ -N-crotonyllysine into histones. *Angew. Chem. Int. Ed.* **51**, 7246–7249 (2012).
191. Schmidt, M. J. & Summerer, D. Red-light-controlled protein-RNA crosslinking with a genetically encoded furan. *Angew. Chem. Int. Ed.* **52**, 4690–4693 (2013).
192. Schmidt, M. J., Borbas, J., Drescher, M. & Summerer, D. A genetically encoded spin label for electron paramagnetic resonance distance measurements. *J. Am. Chem. Soc.* **136**, 1238–1241 (2014).
193. Wang, Y.-S., Russell, W. K., Wang, Z., Wan, W., Dodd, L. E., Pai, P.-J., Russell, D. H. & Liu, W. R. The de novo engineering of pyrrolysyl-tRNA synthetase for genetic incorporation of L-phenylalanine and its derivatives. *Mol Biosyst* **7**, 714–717 (2011).
194. Wang, Y.-S., Fang, X., Wallace, A. L., Wu, B. & Liu, W. R. A rationally designed pyrrolysyl-tRNA synthetase mutant with a broad substrate spectrum. *J. Am. Chem. Soc.* **134**, 2950–2953 (2012).
195. Wang, Y.-S., Fang, X., Chen, H.-Y., Wu, B., Wang, Z. U., Hilty, C. & Liu, W. R. Genetic incorporation of twelve meta-substituted phenylalanine derivatives using a single pyrrolysyl-tRNA synthetase mutant. *ACS Chem. Biol.* **8**, 405–415 (2013).
196. Tuley, A., Wang, Y.-S., Fang, X., Kurra, Y., Rezenom, Y. H. & Liu, W. R. The genetic incorporation of thirteen novel non-canonical amino acids. *Chem. Commun.* **50**, 2673–2675 (2014).
197. Arbely, E., Torres-Kolbus, J., Deiters, A. & Chin, J. W. Photocontrol of tyrosine phosphorylation in mammalian cells via genetic encoding of photocaged tyrosine. *J. Am. Chem. Soc.* **134**, 11912–11915 (2012).
198. Elsässer, S. J., Huang, H., Lewis, P. W., Chin, J. W., Allis, C. D. & Patel, D. J. DAXX envelops a histone H3.3-H4 dimer for H3.3-specific recognition. *Nature* **491**, 560–565 (2012).
199. Arbely, E., Natan, E., Brandt, T., Allen, M. D., Veprintsev, D. B., Robinson, C. V., Chin, J. W., Joerger, A. C. & Fersht, A. R. Acetylation of lysine 120 of p53 endows DNA-binding specificity at effective physiological salt concentration. *Proc. Natl. Acad. Sci.* **108**, 8251–8256 (2011).
200. Umehara, T., Kim, J., Lee, S., Guo, L.-T., Söll, D. & Park, H.-S. N-acetyl lysyl-tRNA synthetases evolved by a CcdB-based selection possess N-acetyl lysine specificity in vitro and in vivo. *FEBS Lett.* **586**, 729–733 (2012).
201. Oppikofer, M., Kueng, S., Martino, F., Soeroes, S., Hancock, S. M., Chin, J. W., Fischle, W. & Gasser, S. M. A dual role of H4K16 acetylation in the establishment of yeast silent chromatin. *EMBO J.* **30**, 2610–2621 (2011).
202. Ko, J.-H., Wang, Y.-S., Nakamura, A., Guo, L.-T., Söll, D. & Umehara, T. Pyrrolysyl-tRNA synthetase variants reveal ancestral aminoacylation function. *FEBS Lett.* **587**, 3243–3248

- (2013).
203. Takimoto, J. K., Dellas, N., Noel, J. P. & Wang, L. Stereochemical basis for engineered pyrrolysyl-tRNA synthetase and the efficient in vivo incorporation of structurally divergent non-native amino acids. *ACS Chem. Biol.* **6**, 733–743 (2011).
204. Lacey, V. K., Louie, G. V., Noel, J. P. & Wang, L. Expanding the library and substrate diversity of the pyrrolysyl-tRNA synthetase to incorporate unnatural amino acids containing conjugated rings. *ChemBioChem* **14**, 2100–2105 (2013).
205. Tharp, J. M., Wang, Y.-S., Lee, Y.-J., Yang, Y. & Liu, W. R. Genetic incorporation of seven ortho-substituted phenylalanine derivatives. *ACS Chem. Biol.* **9**, 884–890 (2014).
206. Johnson, J. A., Lu, Y. Y., Van Deventer, J. A. & Tirrell, D. A. Residue-specific incorporation of non-canonical amino acids into proteins: recent developments and applications. *Curr Opin Chem Biol* **14**, 774–780 (2010).
207. Jing, C. & Cornish, V. W. Chemical tags for labeling proteins inside living cells. *Acc. Chem. Res.* **44**, 784–792 (2011).
208. Hinner, M. J. & Johnsson, K. How to obtain labeled proteins and what to do with them. *Curr. Opin. Biotechnol.* **21**, 766–776 (2010).
209. Prescher, J. A. & Bertozzi, C. R. Chemistry in living systems. *Nat. Chem. Biol.* **1**, 13–21 (2005).
210. Nikić, I., Plass, T., Schraidt, O., Szymański, J., Briggs, J. A. G., Schultz, C. & Lemke, E. A. Minimal tags for rapid dual-color live-cell labeling and super-resolution microscopy. *Angew. Chem. Int. Ed.* **53**, 2245–2249 (2014).
211. Nikić, I., Kang, J. H., Girona, G. E., Aramburu, I. V. & Lemke, E. A. Labeling proteins on live mammalian cells using click chemistry. *Nat Protoc* **10**, 780–791 (2015).
212. Sletten, E. M. & Bertozzi, C. R. Bioorthogonal chemistry: fishing for selectivity in a sea of functionality. *Angew. Chem. Int. Ed.* **48**, 6974–6998 (2009).
213. Lim, R. K. V. & Lin, Q. Bioorthogonal chemistry: recent progress and future directions. *Chem. Commun.* **46**, 1589–1600 (2010).
214. Jacobs, C. L., Yarema, K. J., Mahal, L. K., Nauman, D. A., Charters, N. W. & Bertozzi, C. R. Metabolic labeling of glycoproteins with chemical tags through unnatural sialic acid biosynthesis. *Meth. Enzymol.* **327**, 260–275 (2000).
215. Hang, H. C. & Bertozzi, C. R. Ketone isosteres of 2-N-acetamidoglycans as substrates for metabolic cell surface engineering. *J. Am. Chem. Soc.* **123**, 1242–1243 (2001).
216. Luchansky, S. J., Goon, S. & Bertozzi, C. R. Expanding the diversity of unnatural cell-surface sialic acids. *ChemBioChem* **5**, 371–374 (2004).
217. Sadamoto, R., Niikura, K., Ueda, T., Monde, K., Fukuhara, N. & Nishimura, S.-I. Control of bacteria adhesion by cell-wall engineering. *J. Am. Chem. Soc.* **126**, 3755–3761 (2004).
218. Dirksen, A. & Dawson, P. E. Rapid oxime and hydrazone ligations with aromatic aldehydes for biomolecular labeling. *Bioconjug. Chem.* **19**, 2543–2548 (2008).
219. Zeng, Y., Ramya, T. N. C., Dirksen, A., Dawson, P. E. & Paulson, J. C. High-efficiency labeling of sialylated glycoproteins on living cells. *Nat. Methods* **6**, 207–209 (2009).
220. Rayo, J., Amara, N., Krief, P. & Meijler, M. M. Live cell labeling of native intracellular bacterial receptors using aniline-catalyzed oxime ligation. *J. Am. Chem. Soc.* **133**, 7469–7475 (2011).
221. Nauman, D. A. & Bertozzi, C. R. Kinetic parameters for small-molecule drug delivery by

- covalent cell surface targeting. *Biochim. Biophys. Acta* **1568**, 147–154 (2001).
222. Mahal, L. K., Yarema, K. J. & Bertozzi, C. R. Engineering chemical reactivity on cell surfaces through oligosaccharide biosynthesis. *Science* **276**, 1125–1128 (1997).
223. Griffin, R. J. The medicinal chemistry of the azido group. *Prog Med Chem* **31**, 121–232 (1994).
224. Chang, P. V., Prescher, J. A., Sletten, E. M., Baskin, J. M., Miller, I. A., Agard, N. J., Lo, A. & Bertozzi, C. R. Copper-free click chemistry in living animals. *Proc. Natl. Acad. Sci.* **107**, 1821–1826 (2010).
225. Laughlin, S. T., Baskin, J. M., Amacher, S. L. & Bertozzi, C. R. In vivo imaging of membrane-associated glycans in developing zebrafish. *Science* **320**, 664–667 (2008).
226. Vocadlo, D. J., Hang, H. C., Kim, E.-J., Hanover, J. A. & Bertozzi, C. R. A chemical approach for identifying O-GlcNAc-modified proteins in cells. *Proc. Natl. Acad. Sci.* **100**, 9116–9121 (2003).
227. Ovaa, H., van Swieten, P. F., Kessler, B. M., Leeuwenburgh, M. A., Fiebiger, E., van den Nieuwendijk, A. M. C. H., Galardy, P. J., van der Marel, G. A., Ploegh, H. L. & Overkleeft, H. S. Chemistry in living cells: detection of active proteasomes by a two-step labeling strategy. *Angew. Chem. Int. Ed.* **42**, 3626–3629 (2003).
228. Hangauer, M. J. & Bertozzi, C. R. A FRET-based fluorogenic phosphine for live-cell imaging with the Staudinger ligation. *Angew. Chem. Int. Ed.* **47**, 2394–2397 (2008).
229. Watzke, A., Köhn, M., Gutierrez-Rodriguez, M., Wacker, R., Schröder, H., Breinbauer, R., Kuhlmann, J., Alexandrov, K., Niemeyer, C. M., Goody, R. S. & Waldmann, H. Site-selective protein immobilization by Staudinger ligation. *Angew. Chem. Int. Ed.* **45**, 1408–1412 (2006).
230. Charron, G., Zhang, M. M., Yount, J. S., Wilson, J., Raghavan, A. S., Shamir, E. & Hang, H. C. Robust fluorescent detection of protein fatty-acylation with chemical reporters. *J. Am. Chem. Soc.* **131**, 4967–4975 (2009).
231. Ngo, J. T. & Tirrell, D. A. Noncanonical amino acids in the interrogation of cellular protein synthesis. *Acc. Chem. Res.* **44**, 677–685 (2011).
232. Tsao, M.-L., Tian, F. & Schultz, P. G. Selective Staudinger modification of proteins containing p-azidophenylalanine. *ChemBioChem* **6**, 2147–2149 (2005).
233. Beatty, K. E., Fisk, J. D., Smart, B. P., Lu, Y. Y., Szychowski, J., Hangauer, M. J., Baskin, J. M., Bertozzi, C. R. & Tirrell, D. A. Live-cell imaging of cellular proteins by a strain-promoted azide-alkyne cycloaddition. *ChemBioChem* **11**, 2092–2095 (2010).
234. Hang, H. C., Geutjes, E.-J., Grotenbreg, G., Pollington, A. M., Bijlmakers, M. J. & Ploegh, H. L. Chemical probes for the rapid detection of Fatty-acylated proteins in Mammalian cells. *J. Am. Chem. Soc.* **129**, 2744–2745 (2007).
235. Kho, Y., Kim, S. C., Jiang, C., Barma, D., Kwon, S. W., Cheng, J., Jaunbergs, J., Weinbaum, C., Tamanoi, F., Falck, J. & Zhao, Y. A tagging-via-substrate technology for detection and proteomics of farnesylated proteins. *Proc. Natl. Acad. Sci.* **101**, 12479–12484 (2004).
236. Shelbourne, M., Chen, X., Brown, T. & El-Sagheer, A. H. Fast copper-free click DNA ligation by the ring-strain promoted alkyne-azide cycloaddition reaction. *Chem. Commun.* **47**, 6257–6259 (2011).
237. Salic, A. & Mitchison, T. J. A chemical method for fast and sensitive detection of DNA synthesis in vivo. *Proc. Natl. Acad. Sci.* **105**, 2415–2420 (2008).
238. Staudinger, H. & Meyer, J. *New organic compounds of phosphorus. III. Phosphinemethylene*

- derivatives and phosphinimines. (Helv. Chim. Acta, 1919).
239. van Berkel, S. S., van Eldijk, M. B. & van Hest, J. C. M. Staudinger ligation as a method for bioconjugation. *Angew. Chem. Int. Ed.* **50**, 8806–8827 (2011).
240. Saxon, E. & Bertozzi, C. R. Cell surface engineering by a modified Staudinger reaction. *Science* **287**, 2007–2010 (2000).
241. Lin, F. L., Hoyt, H. M., van Halbeek, H., Bergman, R. G. & Bertozzi, C. R. Mechanistic investigation of the staudinger ligation. *J. Am. Chem. Soc.* **127**, 2686–2695 (2005).
242. Köhn, M. & Breinbauer, R. The Staudinger ligation-a gift to chemical biology. *Angew. Chem. Int. Ed.* **43**, 3106–3116 (2004).
243. Saxon, E., Armstrong, J. I. & Bertozzi, C. R. A ‘traceless’ Staudinger ligation for the chemoselective synthesis of amide bonds. *Org. Lett.* **2**, 2141–2143 (2000).
244. Huisgen, R. *Angew. Chem. Int. Ed. Engl.* **2**, 565 (1963). (Angew. Chem, 1963).
245. Himo, F., Lovell, T., Hilgraf, R., Rostovtsev, V. V., Noodleman, L., Sharpless, K. B. & Fokin, V. V. Copper(I)-catalyzed synthesis of azoles. DFT study predicts unprecedented reactivity and intermediates. *J. Am. Chem. Soc.* **127**, 210–216 (2005).
246. Rostovtsev, V. V., Green, L. G., Fokin, V. V. & Sharpless, K. B. A stepwise huisgen cycloaddition process: copper(I)-catalyzed regioselective ‘ligation’ of azides and terminal alkynes. *Angew. Chem. Int. Ed.* **41**, 2596–2599 (2002).
247. Tornøe, C. W., Christensen, C. & Meldal, M. Peptidotriazoles on solid phase: [1,2,3]-triazoles by regiospecific copper(i)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides. *J. Org. Chem.* **67**, 3057–3064 (2002).
248. Kolb, H. C. & Sharpless, K. B. The growing impact of click chemistry on drug discovery. *Drug Discov. Today* **8**, 1128–1137 (2003).
249. Wang, Q., Chan, T. R., Hilgraf, R., Fokin, V. V., Sharpless, K. B. & Finn, M. G. Bioconjugation by copper(I)-catalyzed azide-alkyne [3 + 2] cycloaddition. *J. Am. Chem. Soc.* **125**, 3192–3193 (2003).
250. Worrell, B. T., Malik, J. A. & Fokin, V. V. Direct evidence of a dinuclear copper intermediate in Cu(I)-catalyzed azide-alkyne cycloadditions. *Science* **340**, 457–460 (2013).
251. Kolb, H. C., Finn, M. G. & Sharpless, K. B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew. Chem. Int. Ed.* **40**, 2004–2021 (2001).
252. Link, A. J. & Tirrell, D. A. Cell surface labeling of Escherichia coli via copper(I)-catalyzed [3+2] cycloaddition. *J. Am. Chem. Soc.* **125**, 11164–11165 (2003).
253. Jao, C. Y. & Salic, A. Exploring RNA transcription and turnover in vivo by using click chemistry. *Proc. Natl. Acad. Sci.* **105**, 15779–15784 (2008).
254. Gierlich, J., Burley, G. A., Gramlich, P. M. E., Hammond, D. M. & Carell, T. Click chemistry as a reliable method for the high-density postsynthetic functionalization of alkyne-modified DNA. *Org. Lett.* **8**, 3639–3642 (2006).
255. Seo, T. S., Bai, X., Ruparel, H., Li, Z., Turro, N. J. & Ju, J. Photocleavable fluorescent nucleotides for DNA sequencing on a chip constructed by site-specific coupling chemistry. *Proc. Natl. Acad. Sci.* **101**, 5488–5493 (2004).
256. Speers, A. E., Adam, G. C. & Cravatt, B. F. Activity-based protein profiling in vivo using a copper(i)-catalyzed azide-alkyne [3 + 2] cycloaddition. *J. Am. Chem. Soc.* **125**, 4686–4687 (2003).

257. Wolbers, F., Braak, ter, P., Le Gac, S., Luttge, R., Andersson, H., Vermes, I. & van den Berg, A. Viability study of HL60 cells in contact with commonly used microchip materials. *Electrophoresis* **27**, 5073–5080 (2006).
258. Hong, V., Steinmetz, N. F., Manchester, M. & Finn, M. G. Labeling live cells by copper-catalyzed alkyne–azide click chemistry. *Bioconjug. Chem.* **21**, 1912–1916 (2010).
259. Link, A. J., Vink, M. K. S. & Tirrell, D. A. Presentation and detection of azide functionality in bacterial cell surface proteins. *J. Am. Chem. Soc.* **126**, 10598–10602 (2004).
260. Kennedy, D. C., Lyn, R. K. & Pezacki, J. P. Cellular lipid metabolism is influenced by the coordination environment of copper. *J. Am. Chem. Soc.* **131**, 2444–2445 (2009).
261. Soriano Del Amo, D., Wang, W., Jiang, H., Besanceney, C., Yan, A. C., Levy, M., Liu, Y., Marlow, F. L. & Wu, P. Biocompatible copper(I) catalysts for in vivo imaging of glycans. *J. Am. Chem. Soc.* **132**, 16893–16899 (2010).
262. Hong, V., Presolski, S. I., Ma, C. & Finn, M. G. Analysis and optimization of copper-catalyzed azide-alkyne cycloaddition for bioconjugation. *Angew. Chem. Int. Ed.* **48**, 9879–9883 (2009).
263. Kennedy, D. C., McKay, C. S., Legault, M. C. B., Danielson, D. C., Blake, J. A., Pegoraro, A. F., Stolorow, A., Mester, Z. & Pezacki, J. P. Cellular consequences of copper complexes used to catalyze bioorthogonal click reactions. *J. Am. Chem. Soc.* **133**, 17993–18001 (2011).
264. Soriano Del Amo, D., Wang, W., Jiang, H., Besanceney, C., Yan, A. C., Levy, M., Liu, Y., Marlow, F. L. & Wu, P. Biocompatible copper(I) catalysts for in vivo imaging of glycans. *J. Am. Chem. Soc.* **132**, 16893–16899 (2010).
265. Jiang, H., Zheng, T., Lopez-Aguilar, A., Feng, L., Kopp, F., Marlow, F. L. & Wu, P. Monitoring dynamic glycosylation in vivo using supersensitive click chemistry. *Bioconjug. Chem.* **25**, 698–706 (2014).
266. Clavadetscher, J., Hoffmann, S., Lilienkamp, A., Mackay, L., Yusop, R. M., Rider, S. A., Mullins, J. J. & Bradley, M. Copper Catalysis in Living Systems and In Situ Drug Synthesis. *Angew. Chem. Int. Ed.* **55**, 15662–15666 (2016).
267. Presolski, S. I., Hong, V., Cho, S.-H. & Finn, M. G. Tailored ligand acceleration of the Cu-catalyzed azide-alkyne cycloaddition reaction: practical and mechanistic implications. *J. Am. Chem. Soc.* **132**, 14570–14576 (2010).
268. Rodionov, V. O., Presolski, S. I., Díaz, D. D., Fokin, V. V. & Finn, M. G. Ligand-accelerated Cu-catalyzed azide-alkyne cycloaddition: a mechanistic report. *J. Am. Chem. Soc.* **129**, 12705–12712 (2007).
269. Besanceney-Webler, C., Jiang, H., Zheng, T., Feng, L., Soriano Del Amo, D., Wang, W., Klivansky, L. M., Marlow, F. L., Liu, Y. & Wu, P. Increasing the efficacy of bioorthogonal click reactions for bioconjugation: a comparative study. *Angew. Chem. Int. Ed.* **50**, 8051–8056 (2011).
270. Uttamapinant, C., Tangpeerachaikul, A., Grecian, S., Clarke, S., Singh, U., Slade, P., Gee, K. R. & Ting, A. Y. Fast, cell-compatible click chemistry with copper-chelating azides for biomolecular labeling. *Angew. Chem. Int. Ed.* **51**, 5852–5856 (2012).
271. Alder, K., Stein, G. & Finzenhagen, H. Über das abgestufte Additionsvermögen von ungesättigten Ringsystemen. *European Journal of Organic Chemistry* **485**, 211–222 (1931).
272. Alder, K. & Stein, G. Zur Polymerisation cyclischer Kohlenwasserstoff. IV. Über die stereoisomeren Formen perhydrierter Naphto- und Anthrachinone. *European Journal of Organic Chemistry* **501**, 247–294 (1933).

273. Wittig, G. & Krebs, A. *On the existence of low-membered cycloalkynes. I.* (Chem. Ber, 1961).
274. Codelli, J. A., Baskin, J. M., Agard, N. J. & Bertozzi, C. R. Second-generation difluorinated cyclooctynes for copper-free click chemistry. *J. Am. Chem. Soc.* **130**, 11486–11493 (2008).
275. Ning, X., Guo, J., Wolfert, M. A. & Boons, G.-J. Visualizing metabolically labeled glycoconjugates of living cells by copper-free and fast Huisgen cycloadditions. *Angew. Chem. Int. Ed.* **47**, 2253–2255 (2008).
276. Boger, D. L. Diels-Alder reactions of heterocyclic aza dienes. Scope and applications. *Chemical Reviews* (1986).
277. Thalhammer, F., Wallfahner, U. & Sauer, J. Reaktivität einfacher offenkettiger und cyclischer dienophile bei Diels-Alder-Reaktionen mit inversem Elektronenbedarf. *Tetrahedron Letters* **31**, 6851–6854 (1990).
278. Balcar, J., Chrisam, G., Huber, F. X. & Sauer, J. Reaktivität von Stickstoff-heterocyclen gegenüber Cyclooctin als Dienophil. *Tetrahedron Letters* **24**, 1481–1484 (1983).
279. Knall, A.-C. & Slugovc, C. Inverse electron demand Diels-Alder (iEDDA)-initiated conjugation: a (high) potential click chemistry scheme. *Chem Soc Rev* **42**, 5131–5142 (2013).
280. Taylor, M. T., Blackman, M. L., Dmitrenko, O. & Fox, J. M. Design and synthesis of highly reactive dienophiles for the tetrazine-trans-cyclooctene ligation. *J. Am. Chem. Soc.* **133**, 9646–9649 (2011).
281. Blackman, M. L., Royzen, M. & Fox, J. M. Tetrazine ligation: fast bioconjugation based on inverse-electron-demand Diels-Alder reactivity. *J. Am. Chem. Soc.* **130**, 13518–13519 (2008).
282. Devaraj, N. K., Weissleder, R. & Hilderbrand, S. A. Tetrazine-based cycloadditions: application to pretargeted live cell imaging. *Bioconjug. Chem.* **19**, 2297–2299 (2008).
283. Devaraj, N. K. & Weissleder, R. Biomedical applications of tetrazine cycloadditions. *Acc. Chem. Res.* **44**, 816–827 (2011).
284. Han, H.-S., Devaraj, N. K., Lee, J., Hilderbrand, S. A., Weissleder, R. & Bawendi, M. G. Development of a bioorthogonal and highly efficient conjugation method for quantum dots using tetrazine-norbornene cycloaddition. *J. Am. Chem. Soc.* **132**, 7838–7839 (2010).
285. Selvaraj, R., Liu, S., Hassink, M., Huang, C.-W., Yap, L.-P., Park, R., Fox, J. M., Li, Z. & Conti, P. S. Tetrazine-trans-cyclooctene ligation for the rapid construction of integrin $\alpha v \beta_3$ targeted PET tracer based on a cyclic RGD peptide. *Bioorg. Med. Chem. Lett.* **21**, 5011–5014 (2011).
286. Liu, D. S., Tangpeerachaikul, A., Selvaraj, R., Taylor, M. T., Fox, J. M. & Ting, A. Y. Diels-Alder cycloaddition for fluorophore targeting to specific proteins inside living cells. *J. Am. Chem. Soc.* **134**, 792–795 (2012).
287. Chen, W., Wang, D., Dai, C., Hamelberg, D. & Wang, B. Clicking 1,2,4,5-tetrazine and cyclooctynes with tunable reaction rates. *Chem. Commun.* **48**, 1736–1738 (2012).
288. Devaraj, N. K., Hilderbrand, S., Upadhyay, R., Mazitschek, R. & Weissleder, R. Bioorthogonal turn-on probes for imaging small molecules inside living cells. *Angew. Chem. Int. Ed.* **49**, 2869–2872 (2010).
289. Haun, J. B., Devaraj, N. K., Hilderbrand, S. A., Lee, H. & Weissleder, R. Bioorthogonal chemistry amplifies nanoparticle binding and enhances the sensitivity of cell detection. *Nat Nanotechnol* **5**, 660–665 (2010).
290. Haun, J. B., Devaraj, N. K., Marinelli, B. S., Lee, H. & Weissleder, R. Probing intracellular biomarkers and mediators of cell activation using nanosensors and bioorthogonal

- chemistry. *ACS Nano* **5**, 3204–3213 (2011).
291. Vázquez, A., Dzajak, R., Dračinský, M., Rampmaier, R., Siegl, S. J. & Vrabel, M. Mechanism-Based Fluorogenic trans-Cyclooctene-Tetrazine Cycloaddition. *Angew. Chem. Int. Ed.* **56**, 1334–1337 (2017).
292. Ou, W., Uno, T., Chiu, H.-P., Grünewald, J., Cellitti, S. E., Crossgrove, T., Hao, X., Fan, Q., Quinn, L. L., Patterson, P., Okach, L., Jones, D. H., Lesley, S. A., Brock, A. & Geierstanger, B. H. Site-specific protein modifications through pyrroline-carboxy-lysine residues. *Proc. Natl. Acad. Sci.* **108**, 10437–10442 (2011).
293. Grünewald, J., Jones, D. H., Brock, A., Chiu, H.-P., Bursulaya, B., Ng, K., Vo, T., Patterson, P., Uno, T., Hunt, J., Spraggon, G. & Geierstanger, B. H. Site-specific dual labeling of proteins by using small orthogonal tags at neutral pH. *ChemBioChem* **15**, 1787–1791 (2014).
294. Chiu, H.-P., Grünewald, J., Hao, X., Brock, A., Okach, L., Uno, T. & Geierstanger, B. H. Simultaneous purification and site-specific modification of pyrroline-carboxy-lysine proteins. *ChemBioChem* **13**, 364–366 (2012).
295. Virdee, S., Kapadnis, P. B., Elliott, T., Lang, K., Madrzak, J., Nguyen, D. P., Riechmann, L. & Chin, J. W. Traceless and site-specific ubiquitination of recombinant proteins. *J. Am. Chem. Soc.* **133**, 10708–10711 (2011).
296. Li, J., Lin, S., Wang, J., Jia, S., Yang, M., Hao, Z., Zhang, X. & Chen, P. R. Ligand-free palladium-mediated site-specific protein labeling inside gram-negative bacterial pathogens. *J. Am. Chem. Soc.* **135**, 7330–7338 (2013).

6 Aims of the thesis

The overall goal of this thesis was to extend the potential of the site-specific pyrrolysine incorporation system. We wanted to achieve this by developing new strategies for all involved steps: (Bio-) synthesis of non-canonical amino acids (NCAA), incorporation of novel NCAs and their bioorthogonal conjugation.

First of all, we wanted to incorporate novel, useful Pyl analogs into proteins such as newly discovered post-translational histone modifications. The possibility of introducing these modifications site-specifically into human histone should facilitate studies related to their natural function.

As major part of this thesis, we planned to focus on the studies of Pyl analog-based bioorthogonal protein modification reactions. We were interested in studying the inverse electron demand Diels-Alder cycloaddition between norbornenes and tetrazines and the incorporation and modification of different norbornyl-Pyl-analogs in proteins in further detail. We wanted to develop a novel bioorthogonal norbornene modification reaction using sulfonyl azides and apply it on norbornyl containing peptides and proteins.

As the major goal of this thesis we defined the utilization of the pyrrolysine biosynthetic machinery to generate an unnatural, double-modifiable Pyl variant which can be subsequently incorporated into a protein of interest. To achieve this, we planned to feed the Pyl precursor 3*R*-methyl-D-ornithine to *E. coli* transfected with the necessary Pyl biosynthesis and incorporation gene cluster. The resulting 3*S*-ethynylpyrrolysine containing protein should be modified and analyzed to prove the correct generation, incorporation and modification of this unnatural amino acid.

Finally, we wanted to demonstrate the benefit of a site specifically incorporated NCAA and its specific modification by a pharmaceutically relevant application. Therefore, we intended to apply site-specifically modified *human carbonic anhydrase*-folic acid-conjugates as targeting, pH responsive capping proteins for drug delivery in mesoporous silica-based nanocarriers.

7 Original publications

This thesis is based on the original publications, which are reproduced in the following subsections with the permission of the copyright holders.

The contributions to each original publication are summarized in section 8.

Relevant parts of the supplementary information to which I mainly contributed are shown in section 9. These supplementary information subsections are only shown for the three original publications in which I am listed as first author (original publications reprinted in sections 7.1, 7.4 and 7.5).

7.1 Synthesis of ϵ -N-propionyl-, ϵ -N-butyryl-, and ϵ -N-crotonyl-lysine containing histone H₃ using the pyrrolysine system

ChemComm

RSCPublishing

COMMUNICATION

View Article Online
View Journal | View IssueSynthesis of ϵ -*N*-propionyl-, ϵ -*N*-butyryl-, and ϵ -*N*-crotonyl-lysine containing histone H3 using the pyrrolysine system†

Michael J. Gattner, Milan Vrabel and Thomas Carell*

Cite this: *Chem. Commun.*,
2013, **49**, 379Received 29th October 2012,
Accepted 14th November 2012

DOI: 10.1039/c2cc37836a

www.rsc.org/chemcomm

Recently new lysine modifications were detected in histones and other proteins. Using the pyrrolysine amber suppression system we genetically inserted three of the new amino acids ϵ -*N*-propionyl-, ϵ -*N*-butyryl-, and ϵ -*N*-crotonyl-lysine site specifically into histone H3. The lysine at position 9 (H3 K9), which is known to be highly modified in chromatin, was replaced by these unnatural amino acids.

The histone code is based on the post-translational modification of critical amino acid residues in different histones. Among these, lysine acetylations and methylations are the most abundant and affect the transcriptional status of the genes associated with the corresponding histones.^{1–3} The role of these modifications is sequence dependent. Typically acetylation is associated with transcriptionally active genes, while methylation induces transcriptional silencing.^{4,5} Recently Zhao and co-workers discovered a number of new modified amino acids in histones.^{6–8} Some of these were also detected in proteins other than histones, raising the possibility that they are of more widespread importance.^{9–11} The new post-translational modifications (PTMs) are acylated derivatives of lysine at the ϵ -amino position. The acylation partners are propionic acid, butyric acid, malonic acid, succinic acid, crotonic acid or fatty acids. A common characteristic of these compounds is that they are key metabolic intermediates that typically exist as CoA activated thioester species in the cell.¹² It is currently not clear how these newly discovered PTMs are biosynthetically established within the histones and we do not fully understand if and how they influence genetic processes.

So far no specific deacylases for ϵ -*N*-propionyl- or ϵ -*N*-butyryl-lysine have been identified. However Sirt5, a member of NAD-dependent sirtuins is able to specifically deacylate ϵ -*N*-

malonyl- and ϵ -*N*-succinyllysine.¹¹ Recently it was discovered that HDAC3 exhibits decrotonylase activity *in vitro*¹³ and it was found that lysine crotonylation activates genes even in a globally repressive environment.^{7,14}

In order to enable investigation of the new lysine derivatives in histones it is essential to generate histones that contain these amino acids site specifically.¹⁵ In this direction semi-synthetic chemical ligation based methods^{16,17} were utilized and chemical methods were employed to generate acetyllysine (Kac) and methylated-lysines or derivatives thereof.^{18–21} Chin and co-workers reported the introduction of Kac into H3 K56²² using the pyrrolysine system. This system was also employed for the synthesis of monomethyl- and dimethyl-lysine containing histones.^{23,24} Schultz *et al.* recently described the synthesis of histone H2B containing ϵ -*N*-crotonyllysine at position K11 using an evolved pyrrolysyl-tRNA synthetase from *Methanosarcina barkeri*.²⁵

Here we show that the pyrrolysine system can be used to insert ϵ -*N*-propionyl-(Kpr, 1), ϵ -*N*-butyryl-(Kbu, 2), and ϵ -*N*-crotonyl-lysine (Kcr, 3) into histones at critical positions such as H3 K9 using the wild type pyrrolysyl-tRNA synthetase. The synthesis of the three modified amino acids 1–3 is shown in Scheme 1. All three amino acids were obtained using the depicted procedure in multi-gram quantities starting from commercially available materials without the need for chromatographic purification (for synthetic details see ESI†).

The amber suppression technology was used extensively for the incorporation of various unnatural amino acids into proteins.²⁶ Recently we reported the development of an *E. coli* based assay that allows examination of how efficiently the pyrrolysine system from *Methanosarcina mazei* works.²⁷ *E. coli* cells containing a plasmid with an YFP reporter gene bearing an amber stop codon at position K114 were generated. The plasmid also contains the genes of the pyrrolysine tRNA (*pylT*) and the corresponding synthetase (*pylS*) or of *pylS* mutants, respectively, that were previously shown to improve the amber suppression efficiency of bulkier pyrrolysine analogues.^{28,29} Upon addition of an unnatural amino acid to the expression medium that is accepted by the tRNA-pyrrolysyl-tRNA synthetase pair (tRNA^{Py1}-PylRS), full-length fluorescent YFP is produced.

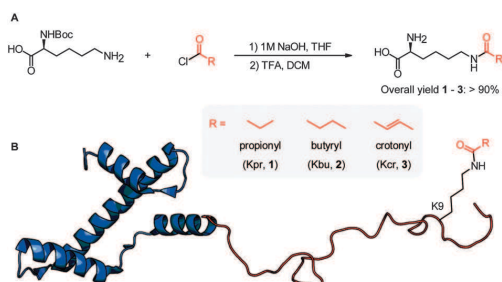
Center for Integrated Protein Science at the Department of Chemistry, Ludwig-Maximilians-Universität, Munich, Butenandtstraße 5-13, D-81377, Munich, Germany. E-mail: Thomas.Carell@cup.uni-muenchen.de;
Fax: +49 (0)89 2180 77756; Tel: +49 (0)89 2180 77750

† Electronic supplementary information (ESI) available: Synthesis of compounds 1–3, preparation of expression vectors, protein expression and purification procedures, HPLC-MS/MS and HPLC-MS data, western blot analysis. See DOI: 10.1039/c2cc37836a

Communication

View Article Online

ChemComm



Scheme 1 (A) Synthesis of the three amino acids ϵ -N-propionyl-(Kpr, **1**), ϵ -N-butryl-(Kbu, **2**), and ϵ -N-crotonyl-Lysine (Kcr, **3**): (1) N_ϵ -Boc-Lys-OH, propionyl-, butyryl or crotonyl chloride, 1 M NaOH/THF = 1/1, 0 °C to r.t., overnight; (2) DCM/TFA = 5/1, r.t., 1.5 h. (B) Depiction of histone H3 and of the position K9 in the tail where the modifications were incorporated (protein representation generated from PDB 1KX5).

Testing the three amino acids Kpr, Kbu and Kcr with this reporter system, the best results were obtained in the presence of wild type PylRS (see ESI†).

To further determine the efficiency of the incorporation we transformed *E. coli* cells with a plasmid containing the gene of thioredoxin A (trxA) with a C-terminal Strep-Tag and an amber codon at position N56 as a model protein. A second plasmid carrying the genes *pylT* and *pylS* was transformed. The cells were grown in the presence of 5 mM Kcr and the protein expression was induced by the addition of 1 mM IPTG at an OD_{600} of 0.6. After an overnight expression at 30 °C the cells were lysed and purified by Strep-Tag affinity chromatography. We obtained 2 mg of the Kcr containing TrxA from 1 L cell culture. The purified intact protein was analyzed by mass spectrometry. The observed protein molecular weight of 13 014 Da (calculated: 13 016 Da) confirmed the presence of the Kcr moiety in the protein. In addition tryptic digest was performed followed by HPLC-MS/MS analysis, which further showed that the modification was in the desired position within the protein structure (see ESI†).

We next investigated if the system is able to provide histone proteins containing one of the three modified lysine amino acids. To this end we used a first plasmid containing *pylS* and three copies of *pylT* under the control of a constitutive glutamine promoter. A second plasmid was generated containing an N-terminally His-tagged H3 gene with the amber stop codon at position K9 under the control of an IPTG-inducible T7 promoter. Both plasmids were transformed into *E. coli* BL21(DE3). We individually added the three amino acids Kpr, Kbu, and Kcr to the medium at concentrations of 5 mM for Kbu and Kcr and 10 mM for Kpr. The *E. coli* cells were grown until an OD_{600} of 0.6 and then gene expression was induced by the addition of 1 mM IPTG. After 16 h at 30 °C we harvested and lysed the cells and isolated the H3 proteins under denaturing conditions (7 M Urea) using Ni-NTA column chromatography. The purified histones containing the modifications in position 9 were refolded using standard procedures (see ESI†). In all three cases we isolated 0.5–1.0 mg of modified histone H3 per L of cell culture (for comparison we obtained 5 mg of unmodified histone H3).

In order to prove the presence of the unnatural lysine amino acids and in particular to investigate potential deacylation processes during protein expression and purification we performed HPLC-MS/MS studies. When we analyzed the tryptic digests of the H3 proteins by HPLC-MS/MS, peptides containing K9 (or the corresponding modifications) were not detected due to the position of this amino acid in a lysine rich sequence giving rise to polar and in particular too short peptides for subsequent analysis. We hence modified a method first described by Hunt *et al.*³⁰ The H3 proteins containing Kbu and Kcr were reacted with propionic acid anhydride (**4**) directly in the gel. For H3 proteins containing the expected Kpr we used butyric acid anhydride (**5**) instead. This chemical treatment blocks all free lysine ϵ -amino groups leading to longer and more hydrophobic tryptic peptides (*e.g.* KSTGGKAPR instead of STGGK for the K9 containing peptide). After in-gel tryptic digest we repeated the chemical treatment to cap also the N-terminal amino groups of the generated peptides. The peptides were then eluted from the gel and analyzed by HPLC-MS/MS. The obtained data are depicted in Fig. 1 (for MS/MS data see ESI†). For wild type histone H3, the positions K9 and K14 are either propionyl- (K9Kpr K14Kpr, as shown in Fig. 1) or butyryl-capped. The MS analyses of the corresponding modified histone H3 peptides gave the correct mass spectra for K9Kpr K14Kbu, K9Kbu K14Kpr and K9Kcr K14pr, respectively, showing the correct incorporation of the three lysine derivatives. We noted that the MS-analyzed peptides were contaminated in all three cases with

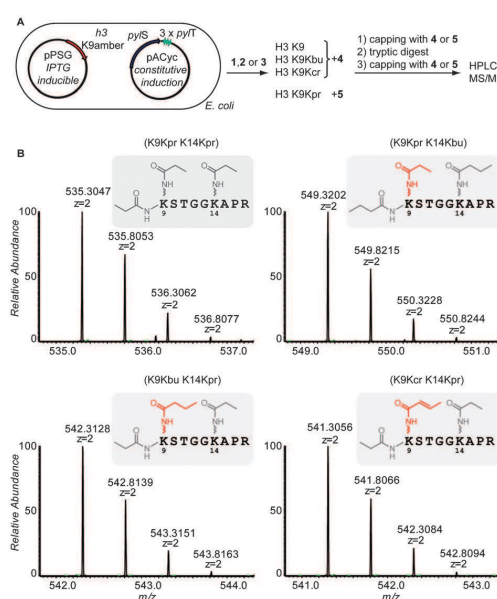


Fig. 1 (A) Schematic overview of the workflow: expression of the four histone H3 variants followed by in-gel capping of the purified proteins with propionic anhydride (**4**) or butyric anhydride (**5**), tryptic digestion, repeated in-gel capping and HPLC-MS/MS analysis. (B) MS spectra and illustrations of the capped K9 containing tryptic peptides (MS/MS spectra are shown in ESI†). The capping modifications are shown in grey while the incorporated PTMs are shown in red.

View Article Online

Communication

ChemComm

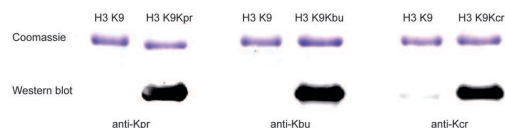


Fig. 2 Western blot analysis of the three synthetic histone variants H3 K9Kpr, H3 K9Kbu and H3 K9Kcr compared to unmodified histone H3 using the corresponding anti-Kpr, -Kbu and -Kcr antibodies.

peptides in which the ϵ -N-modified lysines were replaced by lysine. These peptides gave MS-signals for all-propionylated or all-butyrylated peptides, showing that the acyl groups were missing at H3 K9 already before the peptides were analyzed by mass spectrometry. This discovery excludes the possibility that the acyl groups were cleaved during MS-measurement.

We can also exclude that unmodified lysine was inserted during amber suppression by potentially lysine-mischarged tRNA^{Pro}. When we performed the protein expressions in medium lacking the unnatural amino acids, no YFP or histone H3 was produced, showing that successful amber suppression depends on the presence of these amino acid derivatives. We consequently speculate that *E. coli* contains one or more proteins (such as the bacterial sirtuin CobB) able to unspecifically deacylate Kpr- and Kbu-modified histone tails as previously observed for Kac³¹ and Kcr.²⁵ In order to test this hypothesis we repeated the expression experiment in the presence of 20 mM nicotine amide (NAM), a known inhibitor of sirtuins.³² After careful MS analysis we indeed found a strongly reduced amount of peptides containing natural lysine before the capping. We could reduce the deacylation of Kpr, Kbu and Kcr containing histones from about 20% to below 5% after addition of NAM.

To further verify the presence of the incorporated PTMs we performed a western blot analysis using polyclonal antibodies (PTM BioLabs) against the three amino acids Kpr, Kbu and Kcr. As shown in Fig. 2 the desired modifications were selectively detected by the corresponding antibodies while the unmodified histone H3 was not recognized. The antibodies show compromised selectivities in cross labeling experiments (see ESI†).

In summary, we show that by using the pyrrolysine system the newly discovered lysine derivatives ϵ -N-propionyl-, ϵ -N-butyryl-, and ϵ -N-crotonyl-lysine can be inserted into histones (here H3) at critical positions (here K9) using the wild type pyrrolysyl-tRNA synthetase. The observed deacylation in *E. coli* can be reduced by the addition of nicotine amide and thus homogenous proteins bearing these PTMs are now available. Histones containing defined lysine modifications will enable us to further study the biochemistry behind the acylation and deacylation processes in order to learn how the modification chemistry influences gene activity.

This work was supported by the Excellence Cluster CIPSM, the Volkswagen Foundation and the DFG SFB 1032. M.V. thanks the LMU Munich for postdoctoral fellowship.

Notes and references

- 1 E. I. Campos and D. Reinberg, *Annu. Rev. Genet.*, 2009, **43**, 559–599.
- 2 C. T. Walsh, S. Garneau-Tsodikova and G. J. Gatto Jr., *Angew. Chem., Int. Ed.*, 2005, **44**, 7342–7372.
- 3 M. Biel, V. Wascholowski and A. Giannis, *Angew. Chem., Int. Ed.*, 2005, **44**, 3186–3216.
- 4 B. M. Turner, *Cell*, 2002, **111**, 285–291.
- 5 S. Khorasanizadeh, *Cell*, 2004, **116**, 259–272.
- 6 Y. Chen, R. Sprung, Y. Tang, H. Ball, B. Sangras, S. C. Kim, J. R. Falck, J. Peng, W. Gu and Y. Zhao, *Mol. Cell. Proteomics*, 2007, **6**, 812–819.
- 7 M. Tan, H. Luo, S. Lee, F. Jin, J. S. Yang, E. Montellier, T. Buchou, Z. Cheng, S. Rousseaux, N. Rajagopal, Z. Lu, Z. Ye, Q. Zhu, J. Wysocki, Y. Ye, S. Khochbin, B. Ren and Y. Zhao, *Cell*, 2011, **146**, 1016–1028.
- 8 Z. Xie, J. Dai, L. Dai, M. Tan, Z. Cheng, Y. Wu, J. D. Boeke and Y. Zhao, *Mol. Cell. Proteomics*, 2012, **11**, 100–107.
- 9 J. Garrity, J. G. Gardner, W. Hawse, C. Wolberger and J. C. Escalante-Semerena, *J. Biol. Chem.*, 2007, **282**, 30239–30245.
- 10 Z. Zhang, M. Tan, Z. Xie, L. Dai, Y. Chen and Y. Zhao, *Nat. Chem. Biol.*, 2011, **7**, 58–63.
- 11 J. Du, Y. Zhou, X. Su, J. J. Yu, S. Khan, H. Jiang, J. Kim, J. Woo, J. H. Kim, B. H. Choi, B. He, W. Chen, S. Zhang, R. A. Cerione, J. Auwerx, Q. Hao and H. Lin, *Science*, 2011, **334**, 806–809.
- 12 H. Lin, X. Su and B. He, *ACS Chem. Biol.*, 2012, **7**, 947–960.
- 13 A. S. Madsen and C. A. Olsen, *Angew. Chem., Int. Ed.*, 2012, **51**, 9083–9087.
- 14 E. Montellier, S. Rousseaux, Y. Zhao and S. Khochbin, *Bioessays*, 2012, **34**, 187–193.
- 15 C. A. Olsen, *Angew. Chem., Int. Ed.*, 2012, **51**, 3755–3756.
- 16 P. E. Dawson and S. B. Kent, *Annu. Rev. Biochem.*, 2000, **69**, 923–960.
- 17 T. W. Muir, D. Sondhi and P. A. Cole, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 6705–6710.
- 18 R. K. McGinty, J. Kim, C. Chatterjee, R. G. Roeder and T. W. Muir, *Nature*, 2008, **453**, 812–816.
- 19 J. Guo, J. Wang, J. S. Lee and P. G. Schultz, *Angew. Chem., Int. Ed.*, 2008, **47**, 6399–6401.
- 20 M. D. Simon, F. Chu, L. R. Racki, C. C. de la Cruz, A. L. Burlingame, B. Panning, G. J. Narlikar and K. M. Shokat, *Cell*, 2007, **128**, 1003–1012.
- 21 J. M. Chalker, L. Lercher, N. R. Rose, C. J. Schofield and B. G. Davis, *Angew. Chem., Int. Ed.*, 2012, **51**, 1835–1839.
- 22 H. Neumann, S. M. Hancock, R. Buning, A. Routh, L. Chapman, J. Somers, T. Owen-Hughes, J. van Noort, D. Rhodes and J. W. Chin, *Mol. Cell*, 2009, **36**, 153–163.
- 23 D. P. Nguyen, M. M. Garcia Alai, S. Virdee and J. W. Chin, *Chem. Biol.*, 2010, **17**, 1072–1076.
- 24 D. P. Nguyen, M. M. Garcia Alai, P. B. Kapadnis, H. Neumann and J. W. Chin, *J. Am. Chem. Soc.*, 2009, **131**, 14194–14195.
- 25 C. H. Kim, M. Kang, H. J. Kim, A. Chatterjee and P. G. Schultz, *Angew. Chem., Int. Ed.*, 2012, **51**, 7246–7249.
- 26 M. G. Hoesl and N. Budisa, *Curr. Opin. Biotechnol.*, 2012, **23**, 751–757.
- 27 E. Kaya, K. Gutsmedl, M. Vrabel, M. Muller, P. Thumbs and T. Carell, *ChemBioChem*, 2009, **10**, 2858–2861.
- 28 T. Yanagisawa, R. Ishii, R. Fukunaga, T. Kobayashi, K. Sakamoto and S. Yokoyama, *Chem. Biol.*, 2008, **15**, 1187–1197.
- 29 E. Kaya, M. Vrabel, C. Deiml, S. Prill, V. S. Fluxa and T. Carell, *Angew. Chem., Int. Ed.*, 2012, **51**, 4466–4469.
- 30 B. A. Garcia, S. Mollah, B. M. Ueberheide, S. A. Busby, T. L. Muratore, J. Shabanowitz and D. F. Hunt, *Nat. Protocols*, 2007, **2**, 933–938.
- 31 H. Neumann, S. Y. Peak-Chew and J. W. Chin, *Nat. Chem. Biol.*, 2008, **4**, 232–234.
- 32 B. Schwer, J. Bunkenborg, R. O. Verdin, J. S. Andersen and E. Verdin, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 10224–10229.

7.2 Norbornenes in inverse electron-demand Diels-Alder reactions

Norbornenes in Inverse Electron-Demand Diels–Alder Reactions

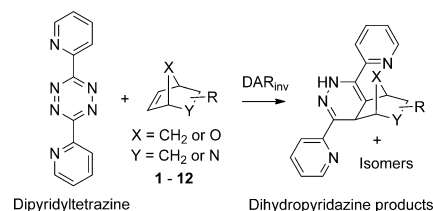
Milan Vrabel,* Patrick Kölle, Korbinian M. Brunner, Michael J. Gattner, Verónica López-Carrillo, Regina de Vivie-Riedle, and Thomas Carell*^[a]

The development of methodologies that enable efficient labeling of biomolecules is an important task in modern organic chemistry. The choice of ideal reaction partners for such a reaction has to fulfil several basic criteria. Along with high reaction rates and high selectivity under physiological conditions that are compatible with the biomolecule, it is also desirable to have good synthetic accessibility to the starting materials. In the last couple of years, several groups reported important developments along these lines that have vastly expanded our ability to functionalize biomolecules.^[1–9] Among the now available methods, the inverse electron-demand Diels–Alder reaction (DAR_{inv}) of tetrazines with strained alkenes has emerged as one of the most promising technologies.^[10,11] The high chemoselectivity and the advantage that the diverse set of functional groups in biomolecules have established the reaction as a truly biorthogonal ligation methodology.^[12–17] In particular the typically high bimolecular rate constants and high yields allow the efficient labeling of fragile proteins given that either the dienophile or the tetrazine is embedded in the biomolecule. However, the synthesis of the reaction partners is quite challenging, which has hindered broad applicability. It is therefore important to analyze the reactivity of different norbornenes and derivatives thereof to develop better accessible starting materials that might even have superior ligation properties to meet the criteria of modern conjugation chemistry.

To optimize the reactivity of the reacting partners in DAR_{inv} it was recently shown that electron-withdrawing substituents at the tetrazine^[18] accelerate the reaction. In addition, the introduction of additional strain into the double- or triple-bond reaction partner, for example, cyclopropane ring fusion to a cyclooctyne or the use of *trans*-cyclooctene, significantly enhance the reaction rate.^[19–22] However, the observed *trans* to *cis* isomerization of cyclooctenes^[23] may lead

to partial deactivation of the reactive species and thus decreases the effective driving force of the reaction. Moreover, cyclooctenes and cyclooctynes are relatively bulky functional groups and may perturb the structure of the target biomolecule. In this regard, cyclopropenes were developed as another type of highly reactive dienophiles, having the advantage of their small size.^[23–25] However, here again the laborious synthesis of this class of compounds may hamper their broader application in bio-labeling experiments. Among other strained dienophiles, norbornenes stand out because they are synthetically easily available, stable, and comparatively small in size. Indeed, others and us used the unusual reactivity of norbornene^[26,27] to label nucleic acids^[28–30] and proteins.^[31–34] We therefore reasoned that we should investigate the parameters that determine the reactivity of norbornenes and of heteroatom-containing norbornene derivatives^[35] with tetrazines to find the best reaction partner in the DAR_{inv} for biomolecule labeling. Here we show that the substitution pattern and in particular the presence of heteroatoms has a dramatic influence on the reactivity (Scheme 1). Our systematic study finally led to the identification of easily available and highly reactive derivatives with significantly enhanced reaction rates and hence ligation properties.

To study how different substituents influence the bimolecular reaction rate of the DAR_{inv}, we first synthesized a series of norbornene derivatives containing either a heteroatom and/or different substituents (Table 1). Compounds **1–12** were synthesized by Diels–Alder or hetero Diels–Alder reactions using cyclopentadiene or furan as the diene, and the corresponding acrylates or in situ formed imines as dienophiles (for details see the Supporting Information). To inves-

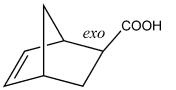

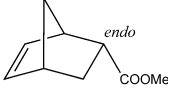
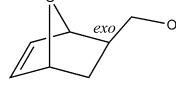
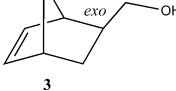
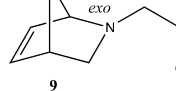
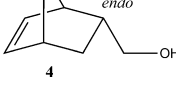
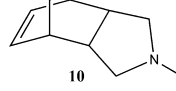

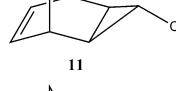
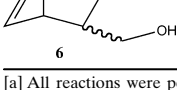
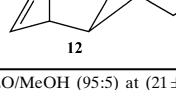


Scheme 1. Schematic representation of the DAR_{inv} between dipyridyltetrazine and norbornene derivatives **1–12** used in this study. The products of the reaction are the corresponding isomers of dihydropyridazines. For substituents R see Table 1.

[a] Dr. M. Vrabel, M.Sc. P. Kölle, M.Sc. K. M. Brunner,
Dipl.-Chem. M. J. Gattner, Dr. V. López-Carrillo,
Prof. Dr. R. de Vivie-Riedle, Prof. Dr. T. Carell
Department of Chemistry
Ludwig-Maximilians-University
Butenandtstrasse 5-13, 81377 Munich (Germany)
Fax: (+49)89 2180 77756
E-mail: milan.vrabel@cup.uni-muenchen.de
thomas.carell@cup.uni-muenchen.de

Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/chem.201301838>.

Table 1. Second-order rate constants for DAR_{inv} of norbornene derivatives **1–12** with dipyridyltetrazine.

Norbornene	k [$\text{M}^{-1}\text{s}^{-1}$] ^[a]	Norbornene	k [$\text{M}^{-1}\text{s}^{-1}$] ^[a]
	1.3 ± 0.02		0.49 ± 0.004
	0.56 ± 0.005		1.78 ± 0.007
	5.2 ± 0.13		0.12 ± 0.001
	1.76 ± 0.013		0.05 ± 0.002
	0.29 ± 0.033		0.43 ± 0.007
	1.43 ± 0.02		4.55 ± 0.087

[a] All reactions were performed in $\text{H}_2\text{O}/\text{MeOH}$ (95:5) at $(21 \pm 0.1)^\circ\text{C}$. The decay of the absorption of dipyridyltetrazine was measured over time with various concentrations of norbornene. The observed reaction rates were obtained by fitting the data with a single exponential equation ($y = a + be^{-kt}$). The linear dependence of the observed rate constants k' on the concentration of norbornene provided the second-order rate constants k . The measurements were repeated three times for each derivative (for details see the Supporting Information).

tigate the influence of the geometry on the reaction rates, we also measured the kinetics of the *exo* and *endo* norbornenes separately. We used dipyridyltetrazine as the reaction partner, because it is commercially available and structurally close to derivatives typically used in bioconjugation experiments. The rate constants were determined under pseudo first-order conditions using UV/Vis spectroscopy following the decay in the absorption of dipyridyltetrazine unit at 300 nm. All reactions were conducted

at $(21 \pm 0.1)^\circ\text{C}$ using a mixture of water/methanol (95/5) as the solvent. The final concentration of dipyridyltetrazine was $50 \mu\text{M}$. The final concentration of norbornenes varied between 0.5 and 5 mM, corresponding to 10–100 equivalents with respect to the tetrazine. The decay in the absorption of the dipyridyltetrazine at 300 nm was monitored over time (for details see the Supporting Information). A typical diagram showing the profile of the observed UV curves is shown in Figure 1. Pseudo first-order rate constants were obtained by fitting the data with a single exponential equation. The observed rate constants were linearly dependent on the concentration of the norbornene reaction partner in all cases. The second-order rate constants were calculated from the linear dependence of the observed rate constants on the concentration of the norbornene studied. The data are listed in Table 1.

The data show that the introduction of a heteroatom into the bicyclic structure of norbornene (derivatives **7**, **8**, and **9**) has in general a negative effect on the reaction rate. The oxygen atom in position 7 shows a less dramatic influence compared to the nitrogen atom present in azanorbornene **9**. In the case of oxanorbornenes (**7** and **8**) the reaction was approximately three times slower than that with compounds **1** and **3**, which do not contain a heteroatom in the bicyclic ring structure. This slower reactivity of the oxanorbornenes may result from the electron-withdrawing effect of the oxygen, which reduces the electron density of the bicyclic ring system. Evidently the shorter C–O bond (143 pm versus 154 pm), which increases the strain in the bicyclic ring structure is unable to overcompensate the $-I$ -effect. We next investigated the electronic effects in more detail studying how substituents influence the reactivity. Indeed, the presence of an electron-withdrawing substituent as in the derivatives **1**, **2**, **5**, and **7** significantly

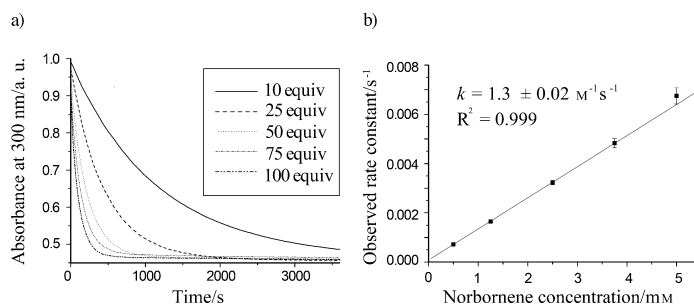


Figure 1. Example of the UV data determined in a typical rate measurement using UV/Vis spectroscopy (shown for derivative **1**): a) graph showing the decrease in UV absorption of the dipyridyltetrazine at 300 nm over time using 10, 25, 50, 75, and 100 equivalents of norbornene **1**; b) plot showing the linear dependence of the observed rate constants on the norbornene concentration. For plots of all norbornene derivatives see the Supporting Information.

reduces the reaction kinetics. In contrast, all derivatives with electron-pushing substituents, such as the alcohols **3**, **4**, **6**, and **8** are 3–4 times faster. Interestingly, the rate acceleration is very high in the *exo*-tricyclo[3.2.1.0^{2,4}]oct-6-ene-3-methanol **12**. In this case, the corresponding alcohol reacted about 10 times faster than the ethyl ester **11**. Because **11** is significantly slower than the *exo* compounds **1** and **3**, again increased ring strain does not account for the higher reactivity of **12**. These results show that additional strain does not increase the reactivity, which is quite surprising. We noted, however, a substantial reactivity difference between the *exo* and the corresponding *endo* norbornene isomers. The large reactivity difference was consistently observed in all examined compounds. The *endo* compounds all react significantly less efficiently. This is particularly impressive for compound **10**. The five-membered ring does not add strain and together with the *endo*-configuration a compound is obtained, which has a strongly reduced reactivity of only $k = (0.05 \pm 0.002) \text{ M}^{-1} \text{ s}^{-1}$. Comparing the rate constants of compound **1** with **2** and that of compound **3** with **4**, an approximately 2.5 fold increase in favor of the *exo* isomer was observed. When we performed the measurements of the two most reactive derivatives **3** and **12** at 37°C, the observed rate constants were $k = (8.0 \pm 0.46) \text{ M}^{-1} \text{ s}^{-1}$ and $k = (6.1 \pm 0.17) \text{ M}^{-1} \text{ s}^{-1}$, respectively, manifesting the potential of these derivatives. Compound **12** has the additional advantage that it contains a mirror symmetry element (C_s symmetry) so that upon reaction only two defined products are formed. The high symmetry of compound **12** together with its superior reactivity makes it the preferred choice for reverse electron-demand Diels–Alder reactions (see Figure S1 in Supporting Information).

To gain further insight into the observed reactivity difference between the *exo* and *endo* isomers, we next performed transition-state calculations of the DAR_{inv} reaction of tetrazine with *exo*-norbornene **3** or *endo*-norbornene **4**, respectively, at the M06-2X/6-311+G(d,p) level of theory. A functional of the M06 family already showed good performance for the computation of barrier heights in other similar inverse electron-demand Diels–Alder reactions.^[19,36,37] The influence of water as solvent was further determined with the polarizable continuum model (PCM).^[38]

The addition of tetrazine to norbornene derivatives **3** and **4** can proceed from the *exo*- or from the *endo*-side, respectively. As a result, four different products can be formed via four different transition states. In all cases the calculated energetic barrier for the *endo*-addition is much higher than for the *exo*-addition (see the Supporting Information). The corresponding reaction leading to the *exo*-addition proceeds for *exo*-norbornene **3** with a lower barrier ($\Delta G^\ddagger = 15.18 \text{ kcal mol}^{-1}$, Figure 2) than for *endo*-norbornene **4** ($\Delta G^\ddagger = 15.83 \text{ kcal mol}^{-1}$). These computations predict that the reaction of tetrazine with **3** should be about three times faster than the same reaction with **4**.^[39] This calculated increase in the rate of the reaction is in excellent agreement with the experimentally observed factor of 2.9 (see Table 1). The explanation for the enhanced stabilization of the transition

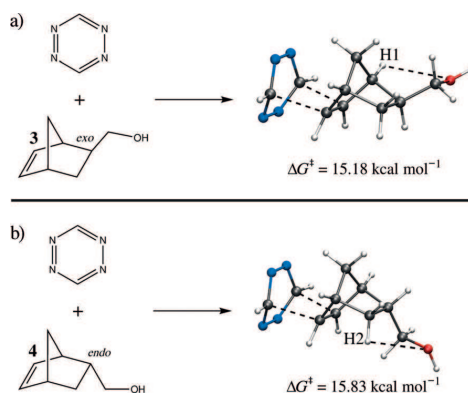


Figure 2. M062X/6-311+G(d,p)-optimized transition-state structures for the inverse electron-demand Diels–Alder reaction of tetrazine with *exo*-norbornene **3** (a) and with *endo*-norbornene **4** (b), both leading to the *exo*-addition of tetrazine. The barrier for the reaction of **3** with tetrazine is 0.65 kcal mol^{−1} lower than the analogous reaction of **4**. The indicated H1–O hydrogen bond (2.66 Å) of **3** is shorter than the corresponding H2–O hydrogen bond (2.85 Å) of **4**.

state for **3** with respect to **4** is the stronger hydrogen bond between H1 and the oxygen atom compared to the H2–O hydrogen bond of **4**.^[40] (Figure 2). This phenomenon was already observed in gas-phase calculations. Further stabilization of the transition state towards **3** occurs due to the interaction of its stronger dipole moment with the solvent water (see the Supporting Information).

In conclusion, we performed a short systematic kinetic study about the reactivity of norbornene derivatives in inverse electron-demand Diels–Alder reactions with tetrazines. We experimentally evaluated the influence of various substituents and of the whole geometry of the norbornene on the reaction rates. The observed data clearly show that norbornenes bearing non-electron-withdrawing substituents in the *exo* position are most favored in the reaction. We performed a computational analysis of the transition states of the DAR_{inv} of tetrazine with *exo* and *endo* norbornene derivatives to rationalize the observed rate difference. The calculations show an excellent agreement with the experimental data and support the observed higher reactivity of the *exo* isomers. We finally conclude that for biomolecular labeling studies one should consider compounds **3** and **12**. The former derivative is synthetically easily available in multi-gram quantities, while the symmetry of compound **12** make this compound the preferred dienophile in inverse electron-demand Diels–Alder reactions with, for example, tetrazines, particularly in cases where a high degree of stereocontrol is important.

Experimental Section

Kinetic measurements: An equal volume of a 0.1 mM solution of dipyr-dyltetrazine in H₂O/MeOH (9:1) and a solution of norbornene (1, 2.5, 5,

7.5 or 10 mM) in H₂O were mixed in an UV-cuvette and immediately inserted into a UV spectrophotometer. The final concentration of tetrazine was 0.05 mM and the final concentration of norbornene 0.5 to 5 mM, corresponding to 10 to 100 fold excess. The decay in the absorption of the tetrazine was followed over 40–60 min in 0.5 or 1 min intervals. The data were fit to a single exponential equation. Each measurement was repeated three times and the observed rate constants k' were plotted against the concentration of norbornene. The second-order constants k were obtained from the slope of the plot.

Acknowledgements

This work was supported by the Excellence Cluster CIPS^M, the Volkswagen Foundation, and the DFG SFB 749 and SFB 1032. M. V. thanks the LMU Munich for a postdoctoral fellowship.

Keywords: inverse electron-demand Diels–Alder reactions • kinetics • norbornene • quantum-chemical calculations

- [1] C. H. Wong, S. C. Zimmerman, *Chem. Commun.* **2013**, 49, 1679–1695.
- [2] P. V. Chang, C. R. Bertozzi, *Chem. Commun.* **2012**, 48, 8864–8879.
- [3] S. S. van Berkel, M. B. van Eldijk, J. C. van Hest, *Angew. Chem.* **2011**, 123, 8968–8989; *Angew. Chem. Int. Ed.* **2011**, 50, 8806–8827.
- [4] R. K. Lim, Q. Lin, *Acc. Chem. Res.* **2011**, 44, 828–839.
- [5] E. Lallana, R. Riguera, E. Fernandez-Megia, *Angew. Chem.* **2011**, 123, 8956–8966; *Angew. Chem. Int. Ed.* **2011**, 50, 8794–8804.
- [6] R. K. Lim, Q. Lin, *Chem. Commun.* **2010**, 46, 1589–1600.
- [7] J. C. Jewett, C. R. Bertozzi, *Chem. Soc. Rev.* **2010**, 39, 1272–1279.
- [8] C. R. Becer, R. Hoogenboom, U. S. Schubert, *Angew. Chem.* **2009**, 121, 4998–5006; *Angew. Chem. Int. Ed.* **2009**, 48, 4900–4908.
- [9] J. M. Baskin, C. R. Bertozzi, *QSAR Comb. Sci.* **2007**, 26, 1211–1219.
- [10] M. L. Blackman, M. Royzen, J. M. Fox, *J. Am. Chem. Soc.* **2008**, 130, 13518–13519.
- [11] N. K. Devaraj, R. Weissleder, S. A. Hilderbrand, *Bioconjugate Chem.* **2008**, 19, 2297–2299.
- [12] K. S. Yang, G. Budin, T. Reiner, C. Vinegoni, R. Weissleder, *Angew. Chem.* **2012**, 124, 6702–6707; *Angew. Chem. Int. Ed.* **2012**, 51, 6598–6603.
- [13] D. S. Liu, A. Tangpeerachaikul, R. Selvaraj, M. T. Taylor, J. M. Fox, A. Y. Ting, *J. Am. Chem. Soc.* **2012**, 134, 792–795.
- [14] R. Rossin, P. R. Verkerk, S. M. van den Bosch, R. C. Vulders, I. Verel, J. Lub, M. S. Robillard, *Angew. Chem.* **2010**, 122, 3447–3450; *Angew. Chem. Int. Ed.* **2010**, 49, 3375–3378.
- [15] J. B. Haun, N. K. Devaraj, S. A. Hilderbrand, H. Lee, R. Weissleder, *Nat. Nanotechnol.* **2010**, 5, 660–665.
- [16] H. S. Han, N. K. Devaraj, J. Lee, S. A. Hilderbrand, R. Weissleder, M. G. Bawendi, *J. Am. Chem. Soc.* **2010**, 132, 7838–7839.
- [17] N. K. Devaraj, R. Upadhyay, J. B. Haun, S. A. Hilderbrand, R. Weissleder, *Angew. Chem.* **2009**, 121, 7147–7150; *Angew. Chem. Int. Ed.* **2009**, 48, 7013–7016.
- [18] M. R. Karver, R. Weissleder, S. A. Hilderbrand, *Bioconjugate Chem.* **2011**, 22, 2263–2270.
- [19] M. T. Taylor, M. L. Blackman, O. Dmitrenko, J. M. Fox, *J. Am. Chem. Soc.* **2011**, 133, 9646–9649.
- [20] A. Borrmann, S. Milles, T. Plass, J. Dommerholt, J. M. Verkade, M. Wiessler, C. Schultz, J. C. van Hest, F. L. van Delft, E. A. Lemke, *ChemBioChem* **2012**, 13, 2094–2099.
- [21] W. Chen, D. Wang, C. Dai, D. Hamelberg, B. Wang, *Chem. Commun.* **2012**, 48, 1736–1738.
- [22] K. Lang, L. Davis, S. Wallace, M. Mahesh, D. J. Cox, M. L. Blackman, J. M. Fox, J. W. Chin, *J. Am. Chem. Soc.* **2012**, 134, 10317–10320.
- [23] J. Yang, J. Seckute, C. M. Cole, N. K. Devaraj, *Angew. Chem.* **2012**, 124, 7594–7597; *Angew. Chem. Int. Ed.* **2012**, 51, 7476–7479.
- [24] D. M. Patterson, L. A. Nazarova, B. Xie, D. N. Kamber, J. A. Prescher, *J. Am. Chem. Soc.* **2012**, 134, 18638–18643.
- [25] Z. Yu, Y. Pan, Z. Wang, J. Wang, Q. Lin, *Angew. Chem.* **2012**, 124, 10752–10756; *Angew. Chem. Int. Ed.* **2012**, 51, 10600–10604.
- [26] J. Spanget-Larsen, R. Gleiter, *Tetrahedron Lett.* **1982**, 23, 2435–2438.
- [27] N. G. Rondan, M. N. Paddon-Row, P. Caramella, I. Mareda, J. P. H. Mueller, K. N. Houk, *J. Am. Chem. Soc.* **1982**, 104, 4974–4976.
- [28] K. Gutmiedl, C. T. Wirges, V. Ehmke, T. Carell, *Org. Lett.* **2009**, 11, 2405–2408.
- [29] J. Schoch, M. Wiessler, A. Jaschke, *J. Am. Chem. Soc.* **2010**, 132, 8846–8847.
- [30] J. Schoch, S. Ameta, A. Jaschke, *Chem. Commun.* **2011**, 47, 12536–12537.
- [31] S. S. van Berkel, A. T. Dirks, M. F. Debets, F. L. van Delft, J. J. Cornelissen, R. J. Nolte, F. P. Rutjes, *ChemBioChem* **2007**, 8, 1504–1508.
- [32] E. Kaya, M. Vrabel, C. Deiml, S. Prill, V. S. Fluxa, T. Carell, *Angew. Chem.* **2012**, 124, 4542–4545; *Angew. Chem. Int. Ed.* **2012**, 51, 4466–4469.
- [33] K. Lang, L. Davis, J. Torres-Kolbus, C. Chou, A. Deiters, J. W. Chin, *Nat. Chem.* **2012**, 4, 298–304.
- [34] T. Plass, S. Milles, C. Koehler, J. Szymanski, R. Mueller, M. Wiessler, C. Schultz, E. A. Lemke, *Angew. Chem.* **2012**, 124, 4242–4246; *Angew. Chem. Int. Ed.* **2012**, 51, 4166–4170.
- [35] P. Quadrelli, M. Mella, P. Paganoni, P. Caramella, *Eur. J. Org. Chem.* **2000**, 20, 2613–2620.
- [36] G. O. Jones, V. A. Guner, K. N. Houk, *J. Phys. Chem. A* **2006**, 110, 1216–1224.
- [37] J. Cioslowski, J. Sauer, J. Hetzenegger, T. Karcher, T. Hierstetter, *J. Am. Chem. Soc.* **1993**, 115, 1353–1359.
- [38] G. Scalmani, M. J. Frisch, *J. Chem. Phys.* **2010**, 132, 114110.
- [39] $k_{rel} = k_1/k_2$ with k_1 and k_2 being the individual rate constants of derivatives 3 and 4 was determined according to the equation $\Delta\Delta G^\ddagger = -RT\ln(k_1/k_2)$.
- [40] T. Steiner, *Angew. Chem.* **2002**, 114, 50–80; *Angew. Chem. Int. Ed.* **2002**, 41, 48–76.

Received: May 13, 2013

Published online: September 11, 2013

7.3 Structural insights into incorporation of norbornene amino acids for click modification of proteins

DOI: 10.1002/cbic.201300435

Structural Insights into Incorporation of Norbornene Amino Acids for Click Modification of Proteins

Sabine Schneider,^[a] Michael J. Gattner,^[a] Milan Vrabec,^[a] Veronika Flügel,^[b] Verónica López-Carrillo,^[a] Stefan Prill,^[a] and Thomas Carell^[a]

The development of chemical tools that enable efficient and selective labelling of proteins is of tremendous importance for basic research and for biomedical and biotechnological applications.^[1] One of the most popular techniques that allow precise control over the modification of proteins is based on the incorporation of genetically encoded non-natural amino acids by using bioorthogonal tRNA/tRNA-synthetase pairs. Crucial to this is the incorporation of amino acids with unique side-chain functionalities that can be selectively addressed by a bioorthogonal chemical reaction.^[2] The pyrrolysine system, which in nature is used to incorporate the amino acid pyrrolysine into proteins, is one of the most popular systems for the incorporation of such noncanonical "clickable" amino acids.^[3] In this regard, amino acids containing alkyne,^[4] trans-cyclooctene,^[5] cyclooctyne,^[5,6] and norbornene^[7] side chains have gained significance, because they can be click-functionalised with azides partially under Cu^I catalysis or with tetrazines in Diels–Alder reactions with inverse electron demand. Both wild-type and specifically mutated pyrrolysine synthetases (PylRSs) have been used to load these noncanonical amino acids onto the cognate pyrrolysine tRNA.^[7c,8]

Although crystal structures of wild-type PylRS have been resolved,^[9] there has been no structural study to explain why structurally demanding amino acids with side chains such as norbornene are recognised. This hampers rational optimisation of the pyrrolysine system. Previously we identified and characterised a *Methanosarcina mazei* PylRS triple mutant (norb-PylRS: Y306G, Y384F, I405R) that accepts the *endo*-norbornene-containing Pyl analogue, **1**.^[7c] Here we show that this protein also loads the more reactive^[10] *exo*-norbornene (**2**) and cyclopropyl-norbornene (**3**) derivatives (Figure 1A) onto the cognate tRNA, for subsequent incorporation into the growing peptide chain at the ribosome. We report click chemistry with these novel norbornene derivatives and provide a crystal structure of norb-PylRS in complex with the nonhydrolysable ATP analogue AMP-PNP, and with the norbornene amino acid **1**. The three amino acids (**1–3**) were synthesised by using the cor-

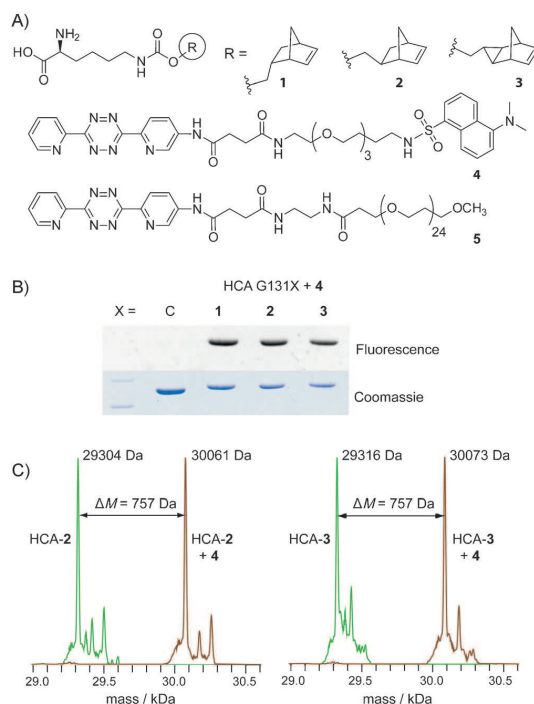


Figure 1. A) Norbornene-containing pyrrolysine analogues **1–3** and tetrazines **4** and **5** (dansyl fluorophore and PEG moiety, respectively). B) SDS-PAGE gel of HCA containing either cysteine or Pyl analogue **1**, **2** or **3** at position 131. All variants were treated with **4**; only HCA samples containing norbornene showed fluorescence. The complete shift of the modified HCA variants compared to the Cys-containing HCA indicates a quantitative modification reaction. C) Intact MS spectra of HCA containing Pyl analogues **2** (HCA-2, left) or **3** (HCA-3, right) at position 131. The deconvoluted MS spectra show unmodified (green) and modified (brown) HCA.

responding norbornene chloroformates and commercially available *N*- α -Boc-protected lysine (see the Supporting Information). Using the pyrrolysine system with norb-PylRS, we incorporated these derivatives into human carbonic anhydrase II (HCA) as a model protein. For protein expression, *Escherichia coli* BL21 (DE3) cells were transformed with two plasmids: one contained the HCA gene with an amber codon at position 131, and a second carried genes coding for *pylT* and *norb-pylRS*. The cells were grown in the presence of 2 mM norbornene amino acid (**1**, **2** or **3**), and protein expression was induced by addition of isopropyl- β -D-thiogalactopyranoside (IPTG; 1 mM) at OD₆₀₀ = 0.9. After 10 h expression at 37 °C, the cells were lysed

[a] Dipl. Chem. M. J. Gattner, Dr. M. Vrabec, Dr. V. López-Carrillo, Dr. S. Prill, Prof. Dr. T. Carell
Department of Chemistry, Ludwig-Maximilians University
Butenadtstrasse 5–13, 81377 Munich (Germany)
E-mail: thomas.carell@cup.uni-muenchen.de

[b] Dr. S. Schneider, MSc V. Flügel
Department of Chemistry, TU Munich
Lichtenbergstrasse 4, 85748 Garching (Germany)
E-mail: sabine.schneider@mytum.de

Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/cbic.201300435>.

and purified by sulfonamide affinity chromatography. From a litre of cell culture, we obtained 9 mg of HCA containing *exo*-norbornene amino acid **2** or 3 mg HCA containing **3**. For comparison, we obtained 9 mg of the protein when using the *endo*-norbornene Pyl analogue **1**. The presence of the noncanonical amino acids at position 131 was confirmed by ESI-MS/MS (see the Supporting Information). The successful incorporation of norbornene derivatives **2** and **3** proves the high promiscuity of the triple mutant norb-PylRS with norbornene substrates.

In order to demonstrate that the norbornene functionalities can be used for site-specific protein modification we next synthesised the tetrazine derivatives **4**, which contains a dansyl fluorophore, and **5**, which can be used for protein PEGylation (for HCA PEGylation see the Supporting Information). We incubated 10 μ M modified HCA (containing **1**, **2** or **3**) in PBS buffer with 5 equivalents of **4** or **5**. After overnight incubation at room temperature, SDS-PAGE gel analysis of the reaction mixtures showed quantitative reactions (Figure 1B). We next confirmed the presence of the expected click products by mass spectrometric analysis of the modified proteins (Figure 1C). In all three cases successful product formation was detected, thus underlining the high efficiency of the norbornene-based protein modification procedure. Most importantly, the cyclopropyl-norbornene (**3**) provides in principle (because of its higher symmetry) only one *exo* and one *endo* product, instead of two of each. This is important because more-complex product mixtures (as obtained with **1** and **2**) can hamper biotechnological applications.

To investigate the molecular basis for the substrate tolerance we determined X-ray crystal structures for the mutant enzyme. Norb-PylRS was crystallised in complex with the ATP analogue AMP-PNP as well as in complex with the adenylated norbornene amino acid (obtained by incubation of the enzyme with **1** and ATP). The structures were solved at 2.35 and 2.45 Å, respectively. (For data processing statistics see Table S1 in the Supporting Information). Norb-PylRS in complex with AMP-PNP superimposes well with the wild-type PylRS structure (PDB ID: 2Q7E), with an RMSD of 0.66 Å. No large structural rearrangements are evident (Figure 2A). In both proteins AMP-PNP binding is the same, with two magnesium ions coordinated by the α/β - and β/γ -phosphates (Figure 2C). In the structure of the

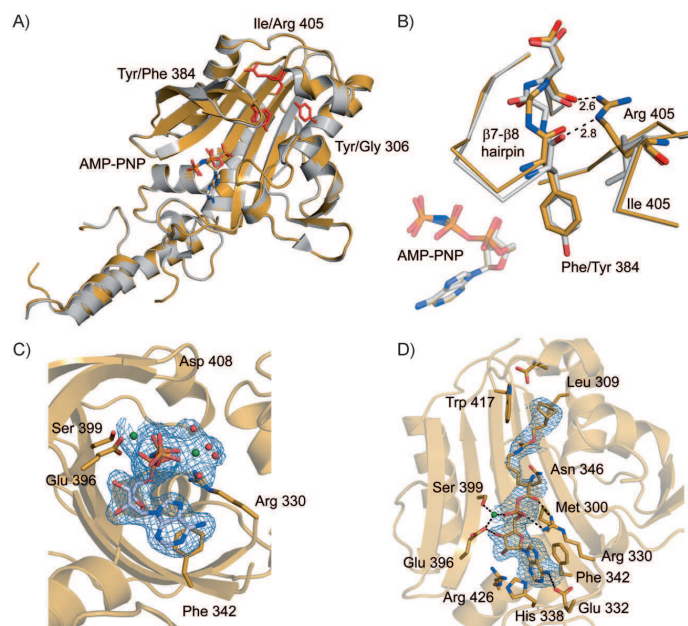


Figure 2. Norb-PylRS triple-mutant structure. A) Overall structure of norb-PylRS in complex with the ATP analogue AMP-PNP (gold) superimposed with the wild-type synthetase (PDB ID: 2Q7E, grey). AMP-PNP shown as a stick model; mutated residues Y306G, Y384F and I405R are in red. B) Ribbon-and-stick representation of the changes in the $\beta 7$ – $\beta 8$ hairpin caused by the I405R mutation. C) and D) Simulated-annealing omit $F_o - DF_c$ electron density map of AMP-PNP (C) and the adenylated norbornene **1** (D) contoured at 3σ . The surrounding amino acid residues are drawn as sticks. Mg^{2+} ions and water are shown as green and red spheres, respectively.

mutant PylRS with AMP-PNP we observed that the Y306G mutation (base of the hydrophobic substrate binding pocket) enlarges the binding site, but does not perturb the backbone structure. The position of Tyr384 in the loop that closes the active side ($\beta 7$ – $\beta 8$ hairpin) is occupied by Phe384. The third mutation (exchange of Ile405 by an arginine) causes a small shift of the $\beta 7$ – $\beta 8$ hairpin as a result of hydrogen-bond formation between Arg405 and the backbone carbonyl of Phe384 and Asp386 (Figure 2B). This loop plays a pivotal role in determining the substrate specificity.^[11] All other parts of the protein backbone are unchanged. We also noticed that in the norb-PylRS:AMP-PNP complex a PEG molecule from the crystallisation solution had entered the active side (Figure S1). In the co-crystals of norb-PylRS with the adenylated norbornene amino acid (**1**), the adenylated norbornene is clearly defined in the electron density (Figure 3D). The norbornene moiety is deeply buried in the enlarged hydrophobic binding pocket (Figure S1). In the wild-type complex, the space occupied by the norbornene head group is taken by the Tyr306 side chain (Figure 3A and B), thus explaining why the Y306G mutation is essential for efficient recognition of the norbornene unit of **1**.^[7c] This mutation enlarges the binding pocket such that the *exo*- and cyclopropyl norbornene units (**2** and **3**) also can be recognised (for **3** see Figure S1). In the wild-type synthetase complex with Pyl, two important interactions can be identified. These are essential for determining the substrate specificity: first, a bifurcat-

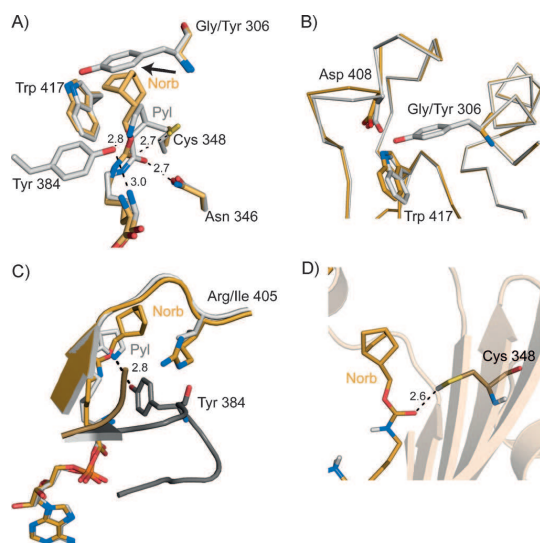


Figure 3. Structural recognition of norbornene amino acid **1** and comparison with the wild-type (wt) enzyme complex. A) Active site with superimposed Pyl and norbornene head groups. The side chain of Tyr306 in the wt enzyme occludes the space occupied by the norbornene functional group in the triple mutant. B) Mutation Y306G does not affect the peptide backbone of the residues at the binding pocket, as shown by the superposition of the PylRS-Norb (gold) and wt (grey) AMP-PNP complexes. C) Closed $\beta 7$ – $\beta 8$ hairpin and Tyr384–Pyl interaction in the wt enzyme and position of Arg405 in the norb-PylRS:norbornene complex. D) Hydrogen bond between the carbamate moiety and Cys348. Cartoon and stick representation of the wt enzyme (grey) and norb-PylRS (gold).

ed hydrogen bond from the tyrosyl residue (Tyr384) to the pyrroline nitrogen as well as to the α -amino group of the substrate (Figure 3C); second, an important H bond between the side chain of Asn346 and the substrate carbonyl-amide linkage.^[8g,9a] In contrast to the crystal structure of the native substrate and *N*-ε-[(cyclopentylloxy)carbonyl-L-lysine] bound to wild-type PylRS, the carbamate linkage of the adenylated *endo*-norbornene (**1**) in complex with the triple mutant norb-PylRS does not form a hydrogen bond with Asn346 (distance 3.45 Å). Instead Cys348 adopts a rotameric conformation in which the cysteine sulfur atom is shifted into an H-bond-donating position (distance ~2.6 Å) to the carbamate oxygen (Figure 3D). Such –SH donor-containing H bonds have been observed to be important for establishing intermolecular interactions;^[12] however, they are generally weaker compared to –NH and –OH donors. The $\beta 7$ – $\beta 8$ hairpin carrying Phe384 is disordered in all structures.^[9a] It was previously proposed that the hairpin functions to protect the unstable pyrrolysyladenylate intermediate when the specific tRNA-Pyl species is in short supply.^[8h,13]

In summary, we present the crystal structure of the pyrrolysyl tRNA synthetase mutant norb-PylRS. The structure unravels the principles that establish the substrate specificity of the norb-PylRS/PylT system for the incorporation of noncanonical amino acids into proteins. Important is the observation that the replacement of Tyr306 by Gly increases the size of the

binding pocket significantly, and therefore enables efficient binding of the sterically demanding norbornene moiety. An unusual H bond between Cys348 and the carbamate oxygen of the amino acid was found to be involved in correct substrate recognition. Additionally, two novel *exo*-norbornene amino acids were successfully incorporated into HCA, thus demonstrating the broad promiscuity of the evolved triple-mutant norb-PylRS. These moieties can be used for quantitative and site-specific fluorescent labelling and PEGylation of proteins, and thus suggests great potential for the norbornene functionality in modern bioorthogonal chemistry.

Experimental Section

General: Chemicals were purchased from Sigma–Aldrich, Fluka, ABCR (Karlsruhe, Germany) or Acros Organics (Fisher Scientific, Nidderau, Germany) and used without further purification. Solvents for silica-gel column chromatography were distilled prior to use. Reactions were monitored on Silica 60 F254 TLC plates (Merck Millipore). Detection was achieved by UV irradiation (254 or 366 nm) and staining with *p*-anisaldehyde in ethanol, CAM or KMnO_4 solutions. Flash column chromatography was performed on Silica 60 (230–400 mesh; Merck Millipore). NMR spectra were recorded on the following spectrometers: Varian Oxford 200 (Agilent Technologies), AC 300 (Bruker), Varian XL 400 and AMX 600 (Bruker). The spectrometers were calibrated by using residual undeuterated solvents as an internal reference. The structures of norbornenes were solved by using 2D NMR techniques. Mass spectra were recorded on the following machines: MAT 95 (EI; Thermo Finnigan), LTQ-FT (ESI-ICR; Thermo Finnigan) and LTQ-Orbitrap XL (Thermo Scientific). Proteins were purified on an ÄKTA purifier system (GE Life Sciences).

Cloning, expression and purification of HCA containing **1, **2** or **3** at position **131**:** The expression vector pACA_HCA G131C (HCA under control of the T7 promoter; ampicillin resistance gene) was a kind donation from Prof. Trauner (Ludwig Maximilians University, Munich). The amber codon (boldface in the following sequence) was introduced to substitute G131 in a site-directed (blunt end) mutagenesis with the 5'-phosphorylated primers 5'-TAG AAA GCT GTG CAG CAA CCT GAT GGA CTG-3' and 5'-AAA ATC CCC ATA TTT GGT GTT CCA GTG AAC CAA G-3' to generate plasmid pACA_HCA G131amber. This plasmid was transformed together with pAcyc_pylRS Norb,^[7c] 3×pylT (genes of the triple mutant of PylRS and three copies of pylT)^[7c] into *E. coli* BL21(DE3) cells (NEB, Norwich, MA). LB medium (1 L) containing chloramphenicol (34 mg L⁻¹), carbenicillin (100 mg L⁻¹) and norbornene amino acid (**1**, **2** or **3**; 2 mM) was inoculated with overnight culture (10 mL). The cells were stirred at 37 °C until OD₆₀₀ = 0.9, then ZnSO₄ (1 mM) and IPTG (0.1 mM) were added to induce the expression of the HCA G131amber gene. After further incubation (10 h, 37 °C) the cells were harvested and stored at –20 °C until further use. The harvested cells were resuspended in wash buffer (Tris (25 mM, pH 8.8), Na₂SO₄ (50 mM), NaClO₄ (50 mM)) and disrupted in a French Press. The supernatant of the centrifuged lysate was subjected to sulfonamide affinity protein purification in an ÄKTA purifier system. A 3 mL column was loaded with *p*-aminomethylbenzenesulfonamide-agarose resin (A0796; Sigma–Aldrich) and equilibrated with wash buffer. After binding (0.75 mL min⁻¹) of the protein solution, the column was washed with seven column volumes of wash buffer. HCA was eluted by low pH (elution buffer: NaOAc (100 mM), NaClO₄ (200 mM, pH 5.6)). The protein-containing fractions were

combined, analysed by SDS-PAGE, dialysed against water and lyophilised. Typical yields of pure HCA protein were 9 mg per L expression medium (for **1** and **2**) and 3 mg per L expression medium (for **3**).

Click of norbornene-containing HCA: Purified HCA containing **1**, **2** or **3** was used for modification with tetrazine derivatives (**4**: dansyl moiety; **5**: PEG moiety). HCA G131C was used as a negative control. The final concentration of the protein was 10 μM in PBS buffer, and tetrazine derivative (50 μM) was added. After overnight incubation at room temperature the samples were analysed by SDS-PAGE or intact mass spectrometry. The SDS-PAGE gel was analysed with an LAS-3000 image reader (Fujifilm) to identify modification of HCA with the fluorophore **4** (excitation 312 nm; Figures S2 and S3).

Intact MS: Purified intact HCA variants were analysed by nano-HPLC-ESI MS. After reaction with tetrazine **4** or **5**, the protein was used directly for HPLC-MS analysis. A self-packed C4 nano-column (ReproSil Gold 300 C4, 3 μm) was used for HPLC separation, and mass spectrometry was performed by an LTQ-Orbitrap XL. Raw mass spectra and deconvoluted spectra are depicted in Figures S7–S12 for all variants.

Norb-PylRS expression, purification and crystallisation: The sequence encoding the catalytic domain (residues 185–454) of the PylRS triple mutant (Y306G, Y384F, I405R; norb-PylRS) from *M. mazei* was PCR amplified with primers 5'-GCG CAT ATG GCA AGT GCC CCC GCA CTT AC-3' and 5'-TTA TGC GGC CGC TTA CAG GTT GGT AG-3' from a plasmid containing the full-length synthetase.^[7c] The fragment was cloned into the expression plasmid pET28a (Novagen/EMD Millipore) in-frame with an N-terminal His₆ tag. Norb-PylRS was expressed in *E. coli* Rosetta DE3 cells (EMD Millipore). As a first purification step the cleared lysate was incubated with Perfect-pro Ni-NTA Superflow resin (5 PRIME, Hamburg, Germany), and eluted with a step-wise imidazole gradient. The eluted protein was concentrated and loaded onto a HiLoad Superdex 200 (16/60) size-exclusion column (GE Healthcare), equilibrated with HEPES (10 mM, pH 7.4), NaCl (300 mM), MgCl₂ (5 mM) and dithiothreitol (1 mM). Fractions containing norb-PylRS were concentrated to 15 mg mL⁻¹. Aliquots were flash-frozen in liquid nitrogen and stored at -80 °C. Crystallisation of norb-PylRS in complex with AMP-PNP was carried out as previously described for the wild-type protein.^[9a] In order to co-crystallise norb-PylRS with the *endo*-norbornene amino acid, the protein was diluted (1 mg mL⁻¹) and incubated for 2 h with *endo*-norbornene sodium salt (2 mM) and ATP (1 mM; Sigma-Aldrich) in protein storage buffer [HEPES (10 mM, pH 7.4), NaCl (300 mM), MgCl₂ (5 mM), dithiothreitol (1 mM)]. The protein was concentrated to 10 mg mL⁻¹ prior to crystallisation. Crystals appeared overnight in sodium acetate (100 mM) and PEG3350 (15–18%). All crystals were cryoprotected with well solution supplemented with ethylene glycol (30%, w/v) before flash-freezing, and then stored in liquid nitrogen until data collection.

Data collection and structure determination: Diffraction data were collected at the synchrotron beam lines PXI (Swiss Light Source, Villigen, Switzerland) and ID23-2 (European Synchrotron Radiation Facility, Grenoble, France). The crystals diffracted X-rays to 2.35 Å (norb-PylRS:AMP-PNP) and 2.45 Å (norb-PylRS:AMP-Norb) spacing, respectively. The data were processed with XDS^[14] and SCALA,^[15] ensuring consistent indexing and choosing the same set of free reflections. The Crystals belong to the same space group as reported previously for the wild type synthetase^[9] (P6₄, unit cell dimensions: $a=b=105$ Å, $c=70$ Å, $\alpha=\beta=90^\circ$, $\gamma=120^\circ$). The structure was solved by molecular replacement by using the norb-PylRS

AMP-PNP wild-type complex (PDB ID:1Q7E) in PHASER.^[16] For the norb-PylRS AMP-Norb the coordinates of the norb-PylRS AMP-PNP were used in a difference Fourier method followed by rigid body refinement in REFMAC.^[17] In order to reduce model bias, all non-protein atoms as well as the loop region around Phe384 were removed from the model prior to molecular replacement, and the temperature factors were reset, followed by simulated annealing in PHENIX.^[18] Clear peaks for AMP-PNP and AMP-Norb were visible in the simulated-annealing omit $F_o - D_F_c$ map (Figure 1). Rounds of model building and refinement were carried out in COOT,^[19] PHENIX and REFMAC. The refinement parameter file for AMP-Norb was generated with prodrgr2^[20] and TLSMD server^[21] was used for TLS refinement. Diffraction data and refinement statistics are summarised in Table S1. All structural figures were prepared with PyMol (Delano Scientific, San Carlos, CA). Atomic coordinates were submitted to the Protein Data Bank (<http://www.ebi.ac.uk/pdbe/>) with the PDB IDs 4BW9 (Pyl-Norb:AMP-PNP) and 4BWA (Pyl-Norb:adenylated norbornene).

Acknowledgements

This work was supported by the Excellence Cluster CIPSM, the Volkswagen Foundation, the DFG (SFB749/A4) and the Fonds der chemischen Industrie (FCI). M.V. thanks the LMU Munich for a postdoctoral fellowship. We thank the beam line scientists at the European Synchrotron Radiation Facility and Swiss Light Source for setting up the beam lines for data collection.

Keywords: click chemistry • norbornenes • protein modifications • pyrrolysine • X-ray structures

- [1] a) Y. Takaoka, A. Ojida, I. Hamachi, *Angew. Chem.* **2013**, *125*, 4182–4200; *Angew. Chem. Int. Ed.* **2013**, *52*, 4088–4106; b) E. M. Sletten, C. R. Bertozzi, *Angew. Chem.* **2009**, *121*, 7108–7133; *Angew. Chem. Int. Ed.* **2009**, *48*, 6974–6998; c) J. Rademann, *Angew. Chem.* **2004**, *116*, 4654–4656; *Angew. Chem. Int. Ed.* **2004**, *43*, 4554–4556; d) A. N. Glazer, *Annu. Rev. Biochem.* **1970**, *39*, 101–130.
- [2] a) L. Davis, J. W. Chin, *Nat. Rev. Mol. Cell Biol.* **2012**, *13*, 168–182; b) C. C. Liu, P. G. Schultz, *Annu. Rev. Biochem.* **2010**, *79*, 413–444; c) J. Xie, P. G. Schultz, *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 775–782.
- [3] a) W.-T. Li, A. Mahapatra, D. G. Longstaff, J. Bechtel, G. Zhao, P. T. Kang, M. K. Chan, J. A. Krzycki, *J. Mol. Biol.* **2009**, *385*, 1156–1164; b) T. Fekner, M. K. Chan, *Curr. Opin. Chem. Biol.* **2011**, *15*, 387–391.
- [4] a) D. P. Nguyen, H. Lusic, H. Neumann, P. B. Kapadnis, A. Deiters, J. W. Chin, *J. Am. Chem. Soc.* **2009**, *131*, 8720–8721; b) E. Kaya, K. Gutsmedl, M. Vrabel, M. Müller, P. Thumbs, T. Carell, *ChemBioChem* **2009**, *10*, 2858–2861; c) T. Fekner, X. Li, M. M. Lee, M. K. Chan, *Angew. Chem.* **2009**, *121*, 1661–1663; *Angew. Chem. Int. Ed.* **2009**, *48*, 1633–1635; d) A. Deiters, T. A. Cropp, M. Mukherji, J. W. Chin, J. C. Anderson, P. G. Schultz, *J. Am. Chem. Soc.* **2003**, *125*, 11782–11783.
- [5] K. Lang, L. Davis, S. Wallace, M. Mahesh, D. J. Cox, M. L. Blackman, J. M. Fox, J. W. Chin, *J. Am. Chem. Soc.* **2012**, *134*, 10317–10320.
- [6] a) A. Borrmann, S. Milles, T. Plass, J. Dommerholt, J. M. Verkade, M. Wiessler, C. Schultz, J. C. M. van Hest, F. L. van Delft, E. A. Lemke, *ChemBioChem* **2012**, *13*, 2094–2099; b) T. Plass, S. Milles, C. Koehler, C. Schultz, E. A. Lemke, *Angew. Chem.* **2011**, *123*, 3964–3967; *Angew. Chem. Int. Ed.* **2011**, *50*, 3878–3881.
- [7] a) T. Plass, S. Milles, C. Koehler, J. Szymanski, R. Mueller, M. Wießler, C. Schultz, E. A. Lemke, *Angew. Chem.* **2012**, *124*, 4242–4246; *Angew. Chem. Int. Ed.* **2012**, *51*, 4166–4170; b) K. Lang, L. Davis, J. Torres-Kolbus, C. Chou, A. Deiters, J. W. Chin, *Nat. Chem.* **2012**, *4*, 298–304; c) E. Kaya, M. Vrabel, C. Deiml, S. Prill, V. S. Fluxa, T. Carell, *Angew. Chem.* **2012**, *124*, 4542–4545; *Angew. Chem. Int. Ed.* **2012**, *51*, 4466–4469.

- [8] a) M. J. Schmidt, D. Summerer, *Angew. Chem.* **2013**, *125*, 4788–4791; *Angew. Chem. Int. Ed.* **2013**, *52*, 4690–4693; b) Z. Yu, Y. Pan, Z. Wang, J. Wang, Q. Lin, *Angew. Chem.* **2012**, *124*, 10752–10756; *Angew. Chem. Int. Ed.* **2012**, *51*, 10600–10604; c) Y.-S. Wang, X. Fang, A. L. Wallace, B. Wu, W. R. Liu, *J. Am. Chem. Soc.* **2012**, *134*, 2950–2953; d) M. Zhang, S. Lin, X. Song, J. Liu, Y. Fu, X. Ge, X. Fu, Z. Chang, P. R. Chen, *Nat. Chem. Biol.* **2011**, *7*, 671–677; e) Y.-S. Wang, W. K. Russell, Z. Wang, W. Wan, L. E. Dodd, P.-J. Pai, D. H. Russell, W. R. Liu, *Mol. Biosyst.* **2011**, *7*, 714–717; f) Z. Hao, Y. Song, S. Lin, M. Yang, Y. Liang, J. Wang, P. R. Chen, *Chem. Commun.* **2011**, *47*, 4502–4504; g) T. Yanagisawa, R. Ishii, R. Fukunaga, T. Kobayashi, K. Sakamoto, S. Yokoyama, *Chem. Biol.* **2008**, *15*, 1187–1197; h) T. Yanagisawa, T. Sumida, R. Ishii, S. Yokoyama, *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2013**, *69*, 5–15; i) T. Mukai, T. Kobayashi, N. Hino, T. Yanagisawa, K. Sakamoto, S. Yokoyama, *Biochem. Biophys. Res. Commun.* **2008**, *371*, 818–822; j) P. R. Chen, D. Groff, J. Guo, W. Ou, S. Cellitti, B. H. Geierstanger, P. G. Schultz, *Angew. Chem.* **2009**, *121*, 4112–4115; *Angew. Chem. Int. Ed.* **2009**, *48*, 4052–4055.
- [9] a) J. M. Kavan, S. Gundllapalli, P. O'Donoghue, M. Englert, D. Söll, T. A. Steitz, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 11268–11273; b) T. Yanagisawa, R. Ishii, R. Fukunaga, O. Nureki, S. Yokoyama, *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **2006**, *62*, 1031–1033.
- [10] M. Vrabel, P. Kölle, K. M. Brunner, M. J. Gattner, V. López-Carrillo, R. de Vries-Riedle, T. Carell, *Chem. Eur. J.* **2013**; DOI: 10.1002/chem.201301838.
- [11] P. O'Donoghue, A. Sethi, C. R. Woese, Z. A. Luthey-Schulten, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 19003–19008.
- [12] P. Zhou, F. Tian, F. Lv, Z. Shang, *Proteins Struct. Funct. Bioinf.* **2009**, *76*, 151–163.
- [13] T. Yanagisawa, R. Ishii, R. Fukunaga, T. Kobayashi, K. Sakamoto, S. Yokoyama, *J. Mol. Biol.* **2008**, *378*, 634–652.
- [14] W. Kabsch, *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66*, 133–144.
- [15] a) M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin, K. S. Wilson, *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2011**, *67*, 235–242; b) P. Evans, *News. Protein Crystallogr.* **1997**, *33*, 22–24.
- [16] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, *J. Appl. Crystallogr.* **2007**, *40*, 658–674.
- [17] G. N. Murshudov, P. Skubák, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long, A. A. Vagin, *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2011**, *67*, 355–367.
- [18] P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66*, 213–221.
- [19] P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66*, 486–501.
- [20] A. W. Schüttelkopf, D. M. Van Aalten, *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2004**, *60*, 1355–1363.
- [21] a) J. Painter, E. A. Merritt, *J. Appl. Crystallogr.* **2006**, *39*, 109–111; b) M. D. Winn, G. N. Murshudov, M. Z. Papiz, *Methods Enzymol.* **2003**, *374*, 300–321.

Received: July 3, 2013

Published online on September 11, 2013

7.4 Sulfonyl azide-mediated norbornene aziridination for orthogonal peptide and protein labeling



ChemComm

COMMUNICATION

View Article Online

View Journal | View Issue



Cite this: *Chem. Commun.*, 2014, 50, 12568

Received 29th May 2014,
Accepted 27th August 2014

DOI: 10.1039/c4cc04117h

www.rsc.org/chemcomm

Sulfonyl azide-mediated norbornene aziridination for orthogonal peptide and protein labeling†

Michael J. Gattner, Michael Ehrlich and Milan Vrabec*‡

We describe a new bioconjugation reaction based on the aziridination of norbornenes using electron-deficient sulfonyl azides. The reaction enables to attach various useful tags to peptides and proteins under mild conditions.

Bioconjugation reactions substantially extend our ability to chemically manipulate proteins. Numerous chemical strategies for the attachment of synthetic molecules to proteins have been developed.¹ Early approaches focused on native functional groups present on endogenous amino acids.² The main drawback of this approach is the lack of specificity since multiple copies of each amino acid are present in the primary protein structure. Unique recognition elements can be introduced into the protein structure to improve the selectivity of the ligation. One possibility is to add a specific amino acid sequence to the target protein that can be recognized by either an appropriate protein modifying enzyme³ or by specific chemical probes.⁴ Another approach uses the biosynthetic machinery of the cell for incorporation of unnatural amino acids containing artificial functional groups.⁵ We and others have used this approach for the incorporation of unnatural amino acids into proteins and have shown that the introduced functionality can be efficiently tagged by orthogonal chemical reactions.⁶ The right choice of the reacting functional groups and the proper ligation technique plays a crucial role here.⁷ Among other suitable functional groups that enable efficient protein labeling various derivatives of cyclooctyne, cyclooctene and cyclopropene have gained special attention in the field.⁸ An extraordinary fast kinetic was observed especially in combination with tetrazines in inverse electron-demand Diels–Alder reactions or in dipolar cycloadditions with

nitrile imines.^{8a,b,e} Unfortunately, the high reactivity of such systems is often accompanied by reduced stability and an increased tendency toward side reactions.⁹ Moreover, synthetic access to these reagents often represents a considerable challenge. It is therefore desirable to develop a methodology that not only enables robust protein labeling but also utilizes easily available starting materials.

Based on our experience using norbornenes as highly reactive compounds in combination with nitrile oxides, nitrile imines and tetrazines¹⁰ we searched for reagents that do react with norbornenes, but also meet the criteria for better synthetic accessibility. Here we show that aziridination of norbornenes using electron-deficient sulfonyl azides represents an excellent balance between reactivity, stability and availability of the starting materials. We demonstrate that this reaction can be used for efficient labeling of norbornene-containing peptides and proteins.

Inspired by the pioneering studies on norbornene aziridination by Franz *et al.*,¹¹ we decided to investigate whether this reaction can proceed in an aqueous environment and can be used for biomolecule labeling. Sulfonyl azides have been previously successfully applied as reagents for detecting thiocarboxylates in bacterial proteomes¹² and as ligation agents for thioacid-containing peptides and proteins.¹³ These studies clearly indicate that sulfonyl azides are compatible with natural systems. To investigate the reactivity of sulfonyl azides with norbornene under aqueous conditions we first performed a model reaction. We reacted sulfonyl azide **1** with norbornene in water/acetonitrile (1:1) at moderate temperature overnight (Scheme 1 and Scheme S1, ESI†). After purification using semipreparative HPLC we obtained the desired aziridine **2** in 78% yield. We also isolated the corresponding sulfonamide **3** (18%) as a by-product formed by the nucleophilic attack of a water molecule on the aziridine and subsequent ring opening. 2D-NMR analysis of **2** showed that only the *exo*-aziridine product was formed in the reaction (see ESI†). The reaction mechanism leading to the formation of the desired aziridine can involve the initial formation of a triazole intermediate followed by the extrusion of nitrogen. Alternatively, the aziridine formation parallels that of epoxidation involving a concerted addition of the azide to the double bond with a concomitant loss of nitrogen.^{11a} No direct

Department of Chemistry, Ludwig-Maximilians-University, Butenandtstrasse 5-13, 81377 Munich, Germany

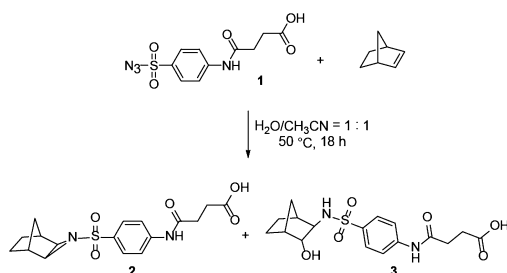
† Electronic supplementary information (ESI) available: Synthesis, experimental details, NMR, mass spectrometry and additional analysis data. See DOI: 10.1039/c4cc04117h

‡ Current address: Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Flemingovo nám. 2, 166 10, Prague, Czech Republic. E-mail: vrabec@uochb.cas.cz; Tel: +420-220183317.

Communication

View Article Online

ChemComm



Scheme 1 Model reaction of norbornene with sulfonyl azide **1**. Yields: 78% of **2** and 18% of **3**.

experimental evidence that would support or disprove any of these mechanisms was observed during our experiments. A more detailed study is required to address this issue.

To gain more insight into the reaction kinetics we measured the bimolecular rate constant of the reaction in water/acetonitrile (9:1) to simulate more relevant conditions required for its use on biomolecules. To ensure sufficient water solubility of the norbornene substrate we used *endo*-5-norbornene-2-methanol in these experiments. The measurements were performed under pseudo first order conditions using an excess of norbornene and were performed in triplicate (for details see ESI†). The second order rate constant was determined to be $k = 1.7 \times 10^{-3} \pm 0.21 \times 10^{-3} \text{ M}^{-1} \text{ s}^{-1}$. This value is comparable to the rate constants of the strain-promoted azide-alkyne cycloaddition reaction of the first generation cyclooctyne derivatives that are commonly used for biomolecule labeling applications.¹⁴ These promising results prompted us to further examine the reaction as a potentially new methodology for modification of biomolecules.

To investigate the stability of sulfonyl azides and products formed in their reaction with norbornenes, we first incubated **1** and the isolated products **2** and **3** in a 50 mM MES buffer at pH 5.5 (MES = 2-(*N*-morpholino)ethanesulfonic acid) and in 50 mM TRIS buffer at pH 8.5 (TRIS = 2-amino-2-hydroxymethylpropane-1,3-diol). HPLC analysis of the mixtures did not show any reaction or decomposition even after a prolonged incubation time (Fig. S6, ESI†). Only slow hydrolysis of aziridine **2** to the corresponding sulfonamide **3** was detected. The observed slow hydrolysis of the originally formed aziridine in water is in fact not an obstacle for biomolecule labeling since the desired tag will stay attached to its target.

Although sulfonyl azides have been previously used in the biological context,^{12,13} to exclude possible side reactions with endogenous amino acids we performed additional experiments. The electron deficient sulfonyl azides could react with nucleophilic groups on proteins (e.g. cysteines or lysines). Our stability study showed that sulfonyl azides were not affected by amino groups since the incubation of compound **1** in TRIS buffer that contains a primary amino group did not show any reaction even at pH 8.5. When we incubated **1** (0.5 mM) with cysteine (50 mM or 2.5 mM) the corresponding sulfonamide was formed as a result of azide reduction (Fig. 1, Fig. S6 and S7, ESI†). No nucleophilic substitution reaction was detected. This reaction again will not interfere with

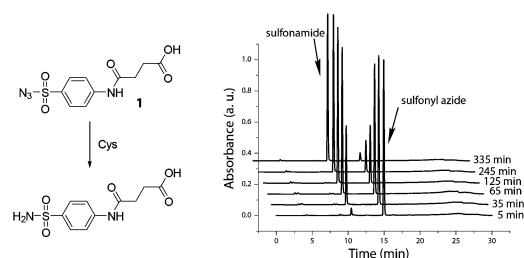


Fig. 1 Reduction of the sulfonyl azide **1** in the presence of cysteine. The new peak (10.3 min) was identified as the corresponding sulfonamide by LC-MS (calc. for $\text{C}_{10}\text{H}_{11}\text{N}_2\text{O}_5\text{S}$ [$\text{M} - \text{H}$] $^-$: 271.0389, found: 271.0393).

protein labeling itself. However, considering the relatively high concentrations of cysteine and glutathione in cells, this side reactivity may hinder the use of this chemistry for *in vivo* applications. To examine a possible reactivity of the aziridine product with other nucleophiles such as amines or thiols we incubated **2** in 50 mM TRIS or in 50 mM cysteine solution. We again observed only slow hydrolysis to form the corresponding sulfonamide **3**. These experiments demonstrate that the aziridine moiety is stable toward these nucleophiles. Also, sulfonyl azides can react with nucleophilic double bonds including indole and *N*-methyl indole.¹⁵ To investigate whether this reaction takes place in tryptophan residues (Trp) we incubated compound **1** (0.5 mM) with Trp (0.5 mM). We did not observe any detectable reaction after two days (Fig. S8, ESI†). These additional experiments indicate that sulfonyl azides can be used for peptide and protein labeling in the presence of these native functional groups.

To evaluate the reaction on biomolecules we next synthesized a norbornene-containing peptide: AFDXKDKPAA, where X = *endo* norbornene-containing amino acid **4**. The peptide was incubated at 50 μM final concentration with sulfonyl azide **1** (2.5 mM) in water/acetonitrile (9:1) and the reaction was followed by MALDI-TOF spectrometry. The analysis of the reaction mixture showed complete conversion of the starting peptide to the aziridinated product within 35 hours (Fig. 2 and Fig. S3, ESI†). The hydrolysis of the aziridine to the corresponding sulfonamide was in this case observed only after prolonged incubation time (72 hours, see Fig. S4, ESI†).

Encouraged by these results we next moved to proteins. Using the pyrrolysine amber suppression system we introduced the *endo* norbornene amino acid **4** into *E. coli* thioredoxin (Trx) and human carbonic anhydrase (HCA) as model proteins.^{10c,16} The norbornene functionality was incorporated at positions Asn65 of Trx (Trx N65X) and His36 of HCA (HCA H36X) (for expression and purification details see the ESI†). We next synthesized sulfonyl azide derivatives **5** and **6** bearing a biotin tag or a dansyl fluorophore to examine the norbornene aziridination on proteins (Fig. 3). We first incubated a 40 μM solution of Trx N65X with 50 equivalents of **5** overnight at 37 °C in 50 mM Tris (pH 7.5). The successful, almost quantitative labeling of Trx N65X with **5** was indicated by a gel shift of the protein band after SDS-PAGE and further confirmed by intact mass spectrometry (Fig. 3A, estimated yield 95%).

In addition, after tryptic digestion and MS analysis we found the expected mass of the peptide LNIDHXPPTAPK containing

ChemComm

View Article Online

Communication

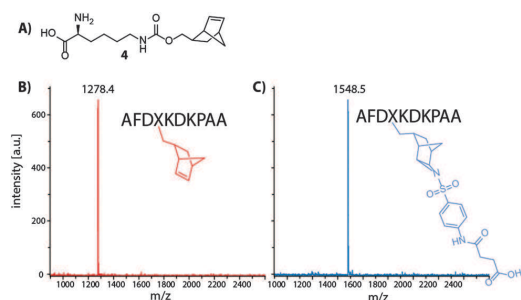


Fig. 2 Aziridination of norbornene-containing peptide AFDXKDKPAA (X = norbornene amino acid **4**) using sulfonamide **1**. (A) Structure of norbornene containing amino acid **4**. (B) MALDI-TOF spectrum of the starting norbornene-containing peptide (calc. mass: $[M - H]^-$: 1279.7 Da). (C) MALDI-TOF spectrum showing the formation of the aziridinated peptide (calc. mass: $[M - H]^-$: 1549.7 Da). $\Delta M_{\text{calc.}} = 270.0$ Da; $\Delta M_{\text{found}} = 270.1$ Da; conditions: 50 μ M peptide, 2.5 mM sulfonamide **1**, $\text{H}_2\text{O}:\text{CH}_3\text{CN} = 9:1$, 37 $^\circ\text{C}$, 35 h.

the desired biotin modification at position X ($[M + 2H]^{2+}_{\text{calc.}} = 1188.0978$, $[M + 2H]^{2+}_{\text{obs.}} = 1188.0919$, $\Delta M = 4$ ppm). To further evaluate the chemistry on proteins we used the wild type HCA (HCA wt) and norbornene containing HCA H36X respectively.

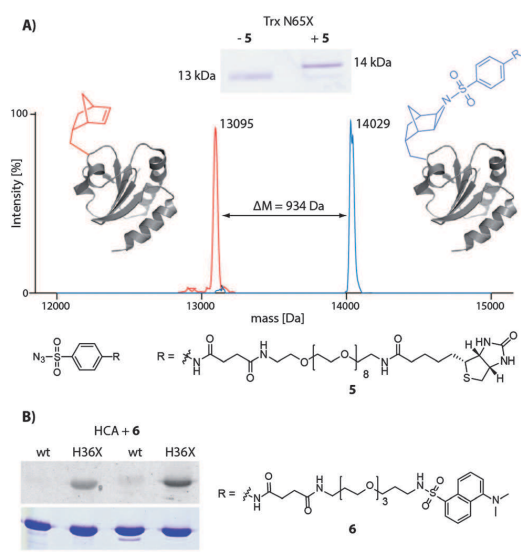


Fig. 3 Aziridination of norbornene-containing proteins. (A) Labeling of thioredoxin N65X (X = norbornene amino acid **4**) using biotin sulfonamide **5**. Coomassie stained gel and overlaid deconvoluted ESI-MS spectra of the norbornene-containing protein before (calc. mass: 13097 Da, obs. mass: 13095 Da) and after reaction with sulfonamide **5** (calc. mass: 14032 Da, obs. mass: 14029 Da, calc. $\Delta M = 935$ Da, obs. $\Delta M = 934$ Da, estimated yield 95%). Depiction of Trx was generated using PDB 2TRX. (B) Fluorescent labeling of HCA H36X (X = norbornene amino acid **4**) using dansyl sulfonamide **6**. SDS-PAGE analysis of the reaction showing specific labeling of the norbornene-containing HCA mutant after 16 h and 4 d. The gel was analyzed by fluorescence detection (upper part) before it was Coomassie stained (lower part).

The proteins (40 μ M final concentration) were incubated with the fluorescent sulfonamide **6** (2 mM, 50 equiv.) in 50 mM MES buffer, pH 5.5 at 37 $^\circ\text{C}$ overnight (Fig. 3B). SDS-PAGE and subsequent fluorescence detection showed selective labeling of the norbornene containing HCA H36X. Only after prolonged incubation time (four days) we observed a weak fluorescent signal in the reaction of the HCA wt protein, which could be due to non-specific labeling. The origin of the non-specific reaction was not clear since our investigations regarding possible side reactions with endogenous amino acids did not show any reaction. Moreover, the sulfonamides were stable under the labeling conditions used (50 mM MES, pH 5.5). Similar non-specific reaction was previously described in the literature, where fluorescent sulfonamides were used for visualization of thiocarboxylates in bacterial proteome.¹² However, in this case also the side reactivity could not be appropriately explained. Nevertheless, the specific reaction with norbornene is much faster and therefore the side reaction can be eliminated by simply removing the excess reagent after the reaction. To verify the presence of the desired fluorescent dansyl tag we digested the protein and analyzed the peptide mixture using HPLC-MS. These data demonstrated that the modification was present in the correct position within the protein structure (peptide QSPVDIDXTAK, X = norbornene amino acid **4** aziridinated by **6**: $[M + 2H]^{2+}_{\text{calc.}} = 1079.5082$, $[M + 2H]^{2+}_{\text{obs.}} = 1079.5039$, $\Delta M = 4$ ppm).

In summary, we show that norbornene aziridination using electron-deficient sulfonamides can be used for orthogonal peptide and protein labeling. The reaction proceeds efficiently under mild conditions, does not require any catalysis and is orthogonal to functional groups of native proteins. Since norbornenes, sulfonamides and derivatives thereof are easily accessible compounds the presented technology constitutes an attractive alternative to currently used bioconjugation techniques especially for *in vitro* applications. Further optimization and evaluation of this chemistry toward its use for *in vivo* peptide and protein labeling is ongoing.

Notes and references

- (a) M. F. Debets, J. C. M. van Hest and F. P. J. T. Rutjes, *Org. Biomol. Chem.*, 2013, **11**, 6439–6455; (b) C. P. Ramil and Q. Lin, *Chem. Commun.*, 2013, **49**, 11007–11022; (c) E. M. Sletten and C. R. Bertozzi, *Angew. Chem., Int. Ed.*, 2009, **48**, 6974–6998.
- (a) E. Basle, N. Joubert and M. Pucheault, *Chem. Biol.*, 2010, **17**, 213–227; (b) G. E. Means and R. E. Feeney, *Bioconjugate Chem.*, 1990, **1**, 2–12.
- (a) M. Rashidian, J. K. Dozier and M. D. Distefano, *Bioconjugate Chem.*, 2013, **24**, 1277–1294; (b) J. M. Chalker, G. J. Bernardes and B. G. Davis, *Acc. Chem. Res.*, 2011, **44**, 730–741.
- (a) T. L. Halo, J. Appelbaum, E. M. Hobert, D. M. Balkin and A. Schepartz, *J. Am. Chem. Soc.*, 2009, **131**, 438–439; (b) J. Zhang, R. E. Campbell, A. Y. Ting and R. Y. Tsien, *Nat. Rev. Mol. Cell Biol.*, 2002, **3**, 906–918.
- (a) J. W. Chin, *Annu. Rev. Biochem.*, 2014, **83**, 379–408; (b) K. Lang and J. W. Chin, *Chem. Rev.*, 2014, **114**, 4764–4806; (c) C. C. Liu and P. G. Schultz, *Annu. Rev. Biochem.*, 2010, **79**, 413–444; (d) W. Wan, J. M. Sharp and W. R. Liu, *Biochim. Biophys. Acta*, 2014, **1844**, 1059–1070; (e) T. S. Young and P. G. Schultz, *J. Biol. Chem.*, 2010, **285**, 11039–11044; (f) Q. Wang, A. R. Parrish and L. Wang, *Chem. Biol.*, 2009, **16**, 323–336.
- (a) T. Fekner, X. Li, M. M. Lee and M. K. Chan, *Angew. Chem., Int. Ed.*, 2009, **48**, 1633–1635; (b) Z. Hao, Y. Song, S. Lin, M. Yang, Y. Liang, J. Wang and P. R. Chen, *Chem. Commun.*, 2011, **47**, 4502–4504; (c) D. P. Nguyen, H. Lusic, H. Neumann, P. B. Kapadnis, A. Deiters and

[View Article Online](#)

Communication

ChemComm

- J. W. Chin, *J. Am. Chem. Soc.*, 2009, **131**, 8720–8721; (d) L. Wang, Z. Zhang, A. Brock and P. G. Schultz, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 56–61; (e) T. Yanagisawa, R. Ishii, R. Fukunaga, T. Kobayashi, K. Sakamoto and S. Yokoyama, *Chem. Biol.*, 2008, **15**, 1187–1197; (f) E. Kaya, K. Gutsmedl, M. Vrabel, M. Müller, P. Thumbs and T. Carell, *ChemBioChem*, 2009, **10**, 2858–2861; (g) N. Li, R. K. V. Lim, S. Edwardraja and Q. Lin, *J. Am. Chem. Soc.*, 2011, **133**, 15316–15319; (h) J. Li, S. Lin, J. Wang, S. Jia, M. Yang, Z. Hao, X. Zhang and P. R. Chen, *J. Am. Chem. Soc.*, 2013, **135**, 7330–7338; (i) A. Dumas, C. D. Spicer, Z. Gao, T. Takehana, Y. A. Lin, T. Yasukohchi and B. G. Davis, *Angew. Chem., Int. Ed.*, 2013, **52**, 3916–3921.
- 7 (a) C. P. Hackenberger and D. Schwarzer, *Angew. Chem., Int. Ed.*, 2008, **47**, 10030–10074; (b) N. Stephanopoulos and M. B. Francis, *Nat. Chem. Biol.*, 2011, **7**, 876–884.
- 8 (a) A. Borrmann, S. Milles, T. Plass, J. Dommerholt, J. M. Verkade, M. Wiessler, C. Schultz, J. C. M. van Hest, F. L. van Delft and E. A. Lemke, *ChemBioChem*, 2012, **13**, 2094–2099; (b) K. Lang, L. Davis, S. Wallace, M. Mahesh, D. J. Cox, M. L. Blackman, J. M. Fox and J. W. Chin, *J. Am. Chem. Soc.*, 2012, **134**, 10317–10320; (c) T. Plass, S. Milles, C. Koehler, C. Schultz and E. A. Lemke, *Angew. Chem., Int. Ed.*, 2011, **50**, 3878–3881; (d) T. Plass, S. Milles, C. Koehler, J. Szymanski, R. Mueller, M. Wiessler, C. Schultz and E. A. Lemke, *Angew. Chem., Int. Ed.*, 2012, **51**, 4166–4170; (e) Z. Yu and Q. Lin, *J. Am. Chem. Soc.*, 2014, **136**, 4153–4156; (f) Z. Yu, Y. Pan, Z. Wang, J. Wang and Q. Lin, *Angew. Chem., Int. Ed.*, 2012, **51**, 10600–10604.
- 9 R. van Geel, G. J. M. Pruijn, F. L. van Delft and W. C. Boelens, *Bioconjugate Chem.*, 2012, **23**, 392–398.
- 10 (a) K. Gutsmedl, C. T. Wirges, V. Ehmke and T. Carell, *Org. Lett.*, 2009, **11**, 2405–2408; (b) M. Vrabel, P. Kolle, K. M. Brunner, M. J. Gattner, V. Lopez-Carrillo, R. de Vivie-Riedle and T. Carell, *Chem. – Eur. J.*, 2013, **19**, 13309–13312; (c) E. Kaya, M. Vrabel, C. Deiml, S. Prill, V. S. Fluxa and T. Carell, *Angew. Chem., Int. Ed.*, 2012, **51**, 4466–4469.
- 11 (a) J. E. Franz, C. Osuch and M. W. Dietrich, *J. Org. Chem.*, 1964, **29**, 2922–2927; (b) J. E. Franz and C. Osuch, *Tetrahedron Lett.*, 1963, **13**, 837–840.
- 12 K. Krishnamoorthy and T. P. Begley, *J. Am. Chem. Soc.*, 2010, **132**, 11608–11612.
- 13 (a) R. Merckx, A. J. Brouwer, D. T. S. Rijkers and R. M. J. Liskamp, *Org. Lett.*, 2005, **7**, 1125–1128; (b) X. Zhang, F. Li, X.-W. Lu and C.-F. Liu, *Bioconjugate Chem.*, 2009, **20**, 197–200.
- 14 (a) N. J. Agard, J. A. Prescher and C. R. Bertozzi, *J. Am. Chem. Soc.*, 2004, **126**, 15046–15047; (b) J. M. Baskin, J. A. Prescher, S. T. Laughlin, N. J. Agard, P. V. Chang, I. A. Miller, A. Lo, J. A. Codelli and C. R. Bertozzi, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 16793–16797.
- 15 (a) A. S. Bailey and J. J. Merer, *J. Chem. Soc. C*, 1966, **15**, 1345–1348; (b) R. E. Harmon, G. Wellman and S. K. Gupta, *J. Heterocycl. Chem.*, 1972, **9**, 1191–1192.
- 16 S. Schneider, M. J. Gattner, M. Vrabel, V. Flügel, V. López-Carrillo, S. Prill and T. Carell, *ChemBioChem*, 2013, **14**, 2114–2118.

7.5 Orchestrating the biosynthesis of an unnatural pyrrolysine amino acid for its direct incorporation into proteins inside living cells

Bioorganic Chemistry

Orchestrating the Biosynthesis of an Unnatural Pyrrolysine Amino Acid for Its Direct Incorporation into Proteins Inside Living Cells

Michael Ehrlich,^[a] Michael J. Gattner,^[a] Bastien Viverge,^[a] Johanna Bretzler,^[a] David Eisen,^[a] Michael Stadlmeier,^[a] Milan Vrabel,^[b] and Thomas Carell^{*,[a]}

Abstract: We here report the construction of an *E. coli* expression system able to manufacture an unnatural amino acid by artificial biosynthesis. This can be orchestrated with incorporation into protein by amber stop codon suppression inside a living cell. In our case an alkyne-bearing pyrrolysine amino acid was biosynthesized and incorporated site-specifically allowing orthogonal double protein labeling.

The incorporation of unnatural amino acids into proteins is a highly topical research field as it allows one to modulate the properties of proteins by chemical means.^[1] Practically, the incorporation of amino acids that are amenable to further chemical modification by orthogonal “click” reaction are interesting for the addition of, for example, polyethylene glycols, toxophores, and fluorophores to proteins for diagnostic or medical applications.^[2] A recently introduced technology in this direction is based on the 22nd genetically encoded amino acid pyrrolysine (Pyl), which is biosynthesized from two L-lysine amino acids by the enzymes PylB, PylC, and PylD (Figure 1 and Figure 1 in the Supporting Information).^[3] The biosynthetic pathway was elucidated by feeding experiments^[4] and structural characterization of the involved enzymes.^[5] Pyl is subsequently loaded onto a special Pyl-tRNA (PylT) with the assistance of a specific aminoacyl-tRNA synthetase (PylRS). PylT delivers Pyl to the ribosome for incorporation into the nascent polypeptide chain in response to the presence of the amber stop codon (UAG) in the mRNA. Therefore the pyrrolysine system reprograms this stop codon for Pyl insertion.

As PylRS is rather promiscuous and can be adopted to bind unnatural amino acids using bio-engineering, the PylRS/PylT pair has enabled the incorporation of a large variety of Pyl de-

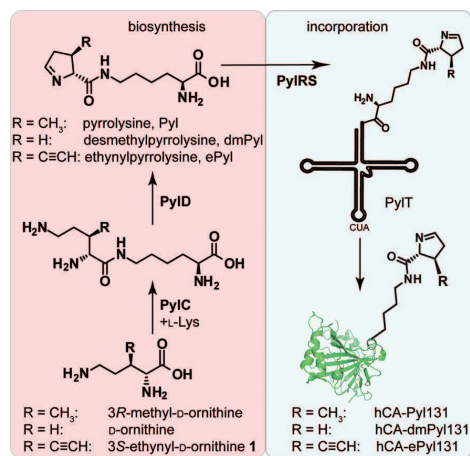


Figure 1. Developed methodology for biosynthesis and incorporation of the novel ethynylpyrrolysine (ePyl). The new amino acid 3S-ethynyl-D-ornithine 1 was converted to ePyl by hijacking the pyrrolysine biosynthesis machinery analogous to the biosynthesis of pyrrolysine (Pyl) via biosynthesis enzymes PylC and PylD. A site-specific incorporation into the protein human carbonic anhydrase 2 (hCA, PDB: 1BCD) by translation with an aminoacyl-tRNA synthetase (PylRS)/Pyl-tRNA (PylT) pair at the amber stop codon was possible.

derivatives featuring additional functional groups, for example, for click modification.^[6] One problem associated with this technology is, however, that the unnatural Pyl amino acids need to be synthesized in rather large quantities to reach the required millimolar concentrations in the culture medium necessary for efficient incorporation. These laborious syntheses can be circumvented with a host organism (*E. coli*) that produces the modified unnatural amino acid by itself.^[7] This in consequence requires the insertion and modulation of the Pyl biosynthetic machinery in the *E. coli* host system.^[4b,c] Thus we planned to study if the Pyl biosynthesis pathway allows in vivo construction and finally incorporation of the unnatural 3S-ethynylpyrrolysine (ePyl) amino acid, which may be amenable to subsequent click modification (Figure 1).

Since Pyl is derived from the PylB product 3R-methyl-D-ornithine and one L-lysine via the action of the enzymes PylC and PylD (see Figure 1 in the Supporting Information), we cloned the genes of both enzymes into a first plasmid (1) (Figure 2). A second plasmid (2) was prepared containing one gene copy of the PylRS aminoacyl-tRNA synthetase gene *pylS* and three

[a] M. Ehrlich,^{*} M. J. Gattner,^{*} B. Viverge, J. Bretzler, Dr. D. Eisen, M. Stadlmeier, Prof. Dr. T. Carell
Center for Integrative Protein Science at the Department of Chemistry
Ludwig-Maximilians-Universität München
Butenandtstrasse 5–13, 81377 Munich (Germany)
E-mail: Thomas.Carell@lmu.de

[b] Dr. M. Vrabel
Institute of Organic Chemistry and Biochemistry
Academy of Sciences of the Czech Republic
Flemingovo n.á. 2, 166 10, Prague (Czech Republic)

^[*] These authors contributed equally to this work.

Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/chem.201500971>.

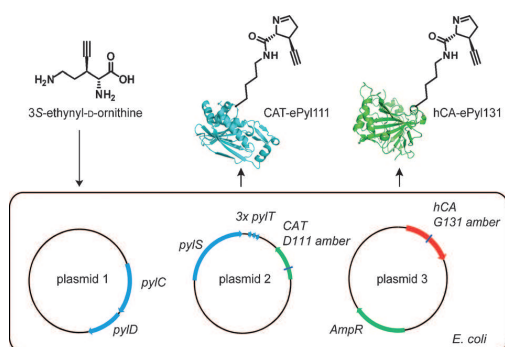


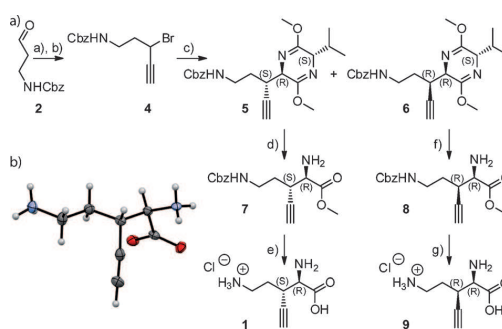
Figure 2. Biosynthetic conversion of 3S-ethynyl-D-ornithine **1** to ethynylpyrrolysine (ePyl) and incorporation into the amber codon bearing proteins chloramphenicol acetyltransferase (CAT) and human carbonic anhydrase 2 (hCA) using an *E. coli* expression system. *E. coli* cells were transformed with following coding plasmids: plasmid 1: *pylC*, *pylD*; plasmid 2: *pylS*, 3x *pylT* and *CAT-A69*amber or *CAT-D111*amber; plasmid 3: *hCA131*amber and an ampicillin resistance gene (PDBs: 1BCD, 1Q23).

copies of the *pylT* gene. A third plasmid (3) was used with the gene for the human carbonic anhydrase 2 (hCA).^[8] This manipulated hCA gene contains an amber stop codon (GGG→TAG) at the amino acid position G131.^[9] All three plasmids were co-transformed into *E. coli* cells. Plasmid 3 was selected on a carbenicillin-containing medium, while plasmid 2 was equipped with a chloramphenicol acetyltransferase (CAT) gene bearing an amber stop codon at either position A69 or D111 upstream of the active site. This plasmid represents the master reporter as it allows growth of *E. coli* cells in the presence of chloramphenicol only if a Pyl derivative is biosynthesized and incorporated in CAT efficiently by stop codon suppression.

For selection of successfully transformed *E. coli* cells we added carbenicillin, chloramphenicol, and initially D-ornithine. This system allowed us to test if D-ornithine is converted to desmethylpyrrolysine (dmPyl) by the action of PylC and PylD followed by final incorporation into proteins via amber suppression.^[4b] Indeed, we observed growing *E. coli* showing that the biosynthesis operates and thus dmPyl is successfully incorporated into CAT.

We next used 3'-substituted D-ornithine derivatives and finally found growing *E. coli* cells in the presence of carbenicillin, chloramphenicol, and 3S-ethynyl-D-ornithine **1**. This demonstrates that 3S-ethynyl-D-ornithine **1** is able to serve as a D-ornithine surrogate during biosynthesis.

The synthesis of the starting material 3S-ethynyl-D-ornithine **1** is conveniently possible as depicted in Scheme 1. We started with Carboxybenzyl (Cbz) protected 3-amino-propionaldehyde **2**, which was allowed to react with ethynyl bromide in a Grignard reaction. The formed alcohol **3** was converted into the bromide **4** via an Appel reaction.^[10] Subsequent substitution of the bromide **4** with the deprotonated S-configured Schöllkopf reagent provided the two pyrazine derivatives with the corresponding SR (**5**) and RR (**6**) configuration in a diastereoisomeric mixture which could be separated using silica column chromatography. We were unable to detect the SS and RS



Scheme 1. a) Synthesis of 3S-ethynyl-D-ornithine **1** and 3R-ethynyl-D-ornithine **9**. Reagents and conditions: a) $\text{HC}\equiv\text{CMgBr}$, THF, -78°C to RT, 2 h; b) PPh_3 , CBr_4 , DCM, 0°C to RT, 5 h; 50% over 2 steps; c) (S)-2,5-dihydro-3,6-dimethoxy-2-isopropylpyrazine, $n\text{BuLi}$, THF, -78°C , 30 min, 72% (SR/RR: 2/5); d) 3 M HCl, MeOH, H_2O , RT, 2 d, 78%; e) 12 M HCl, H_2O , 80°C , 15 h, 91%; f) 3 M HCl, MeOH, H_2O , RT, 2 d, 86%; g) 12 M HCl, H_2O , 80°C , 15 h, 11%; h) X-ray structure of **1**.

products, showing the high stereocontrol provided by the Schöllkopf auxiliary.^[11] Sequential deprotection of both diastereoisomers to the methoxiesters **7** and **8** and subsequent hydrolysis provided the 3S,2R- and 3R,2R-configured ethynylornithines **1** and **9**, respectively. The absolute configuration was verified by a crystal structure of the 3S,2R-ethynylornithine **1** (Scheme 1b).

To investigate how efficiently 3S-ethynyl-D-ornithine **1** serves as a biosynthetic precursor for Pyl derivatives we performed time-dependent growth studies. Growth was observed in the presence of D-ornithine and 3S-ethynyl-D-ornithine **1** but interestingly not with 3R-ethynyl-D-ornithine **9** (see Figure 2 in the Supporting Information). We explain this result with the fact that PylC has a hydrophobic pocket to bind the methyl group of its natural 3R-methyl-D-ornithine substrate that is large enough to accommodate the 3S-ethynyl unit.^[5b] Interestingly, *E. coli* cells show faster growth in the presence of 3S-ethynyl-D-ornithine compared to D-ornithine, indicating that filling of the hydrophobic pocket enhances substrate affinity and hence the efficiency of the biosynthesis (Figure 3 in the Supporting Information).

To prove that the designed *E. coli* cells are indeed able to manufacture 3S-ethynylpyrrolysine (ePyl) for subsequent incorporation into proteins, we next isolated the hCA protein from the *E. coli* cultures (see Figure 4 in the Supporting Information). The protein was conveniently purified in a single chromatographic step using a *p*-aminomethylbenzene-sulfonamide column exploiting the binding affinity to the Zn^{II} ion in the active site of hCA.^[12] When we added D-ornithine in our control experiments to the medium we typically isolated around 6 mg of hCA per L of culture. Upon addition of 3S-ethynyl-D-ornithine **1**, the yield increased dramatically to 75 mg per L indicating, in combination with the results from the growth curves, that the speed of the biosynthesis of ePyl is the determining factor.

Subsequently, the incorporation of ePyl into the hCA enzyme was proven by intact protein mass spectrometric stud-

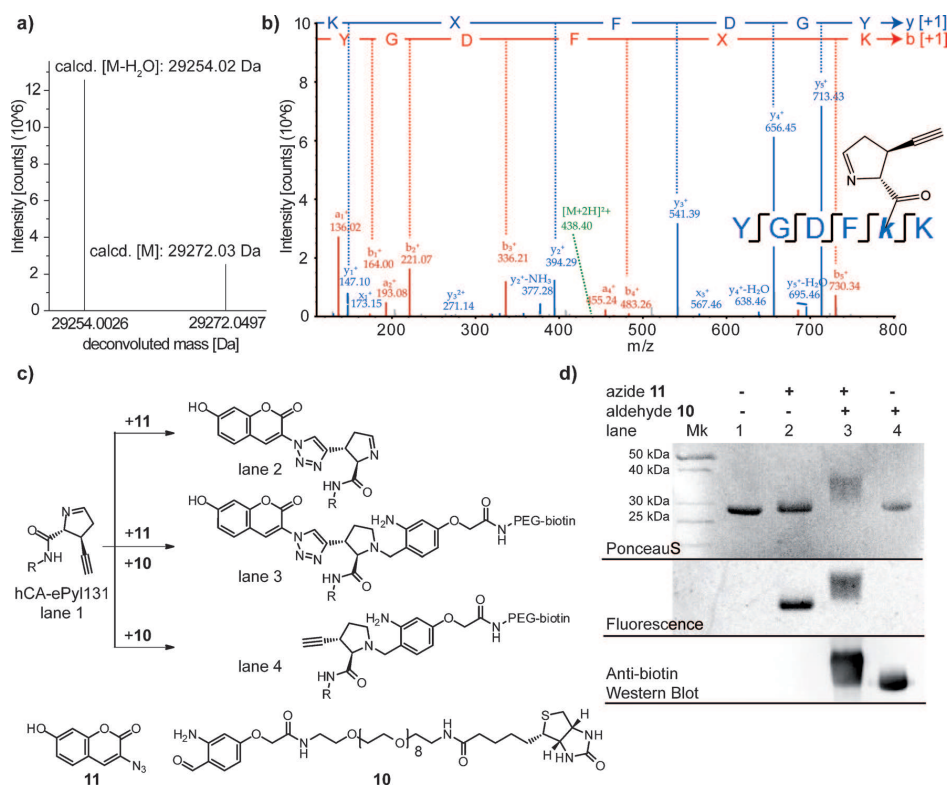


Figure 3. Site-specific incorporation of ePyl in hCA and orthogonal double labeling. a) Deconvoluted intact mass spectrum of hCA-ePyl131. Detected protein species: [M]: 29272.05 Da (calcd.: 29272.03 Da; relative abundance 20%), [M-H₂O]: 29254.00 (calcd.: 29254.02 Da; relative abundance 100%). b) Identification of ePyl peptide fragment after tryptic digestion: hCA sequence coverage of 88.03%. MS/MS spectra of peptide fragment YGDFXK: (X = ePyl). Identified a- and b-ions are marked in red, identified x- and y-ions are marked in blue. Parent ion: [M+2H]²⁺ = 438.71616 (ΔM = -0.01 ppm). c) Scheme of the single and/or double ePyl modification reaction with either azide **11** or the 2-aminobenzaldehyde reagent **10** or using both. d) Labeling of hCA-ePyl131. Protein bands were visualized for all samples after SDS-PAGE by fluorescence measurement. After transfer to a nitrocellulose membrane a PonceauS staining followed by Western Blot using an anti-biotin-HRP conjugate and chemiluminescent readout was performed.

ies using a high-resolution FT-ICR mass spectrometer. Indeed, the isolated hCA protein showed the correct molecular weight for the ePyl-containing hCA-ePyl131 (Figure 3a; see also Figure 5 in the Supporting Information). The position of ePyl incorporation was verified by the mass spectrometric analysis after tryptic digestion on an Orbitrap XL system (LC-MS/MS). This experiment demonstrates that the incorporation proceeded at the desired position (G131→ePyl131, Figure 3b, see also Figure 6 in the Supporting Information). In addition we verified the correct incorporation in CAT-ePyl111 (see Figure 7 in the Supporting Information). As a further control we analyzed the isolated hCA-dmPyl131 protein that was generated in the presence of D-ornithine and observed the correct molecular weight for dmPyl incorporation (G131→dmPyl131; see Figure 8 in the Supporting Information). In summary the data show that our Pyl biosynthesis cluster containing *E. coli* cells produced an hCA protein variant with a G131→ePyl131 mutation.

To demonstrate that the ePyl amino acid can be further manipulated inside a protein we next employed two orthogonal

chemical modification reactions. We first used the chemistry developed by Geierstanger and co-workers^[13] to form a tertiary amine adduct by site-specific reaction at the pyrroline ring structure of Pyl with 2-aminobenzaldehyde (ABA) derivatives. We treated to this end the ePyl-containing hCA with a biotinylated 2-aminobenzaldehyde derivative **10** shown in Figure 3c. The reagent was added to the hCA-ePyl131 protein solution and the protein was analyzed after 16 h reaction at room temperature by gel electrophoresis. Full conversion of the hCA-ePyl131 protein was observed based on migration differences of the labeled protein. The biotinylation of hCA-ePyl131 was verified in a Western blot using an anti-biotin antibody (Figure 3d).

Next, we studied the ability to click-react the alkyne unit by Cu^I-catalyzed reaction (CuAAC) with 7-hydroxycoumarin azide (**11**), which turns fluorescent after cycloaddition.^[14] Indeed, after treating the protein at room temperature for 16 h at pH 7.4 the product protein showed the expected blue fluorescence (Figure 3d). Tryptic digestion and LC-MS/MS analysis al-

lowed us to unambiguously identify the coumarin-labeled protein (see Figure 9 in the Supporting Information).

Finally, we investigated if the ePyl amino acid can be used for double labeling. To this end the hCA-ePyl131 protein was first treated with azide **11** (2 h, RT, pH 7.4) and subsequently reacted with the biotinylated 2-aminobenzaldehyde **10** (14 h, RT, pH 7.4). The obtained product protein showed blue fluorescence and presence of the inserted biotin label was proven by immunostaining with the anti-biotin antibody. Further LC-MS/MS analysis provided additional support for the successful double labeling of the hCA protein (Figure 3 d, see Figure 10 in the Supporting Information).

In summary we generated an *E. coli* expressions system which is able to biosynthesize an unnatural amino acid. This biosynthesis can be orchestrated with methods that allow the incorporation of the resulting unnatural amino acid into proteins by amber suppression. In this particular case we inserted an alkyne-modified pyrrolysine amino acid which is amenable to two different orthogonal click reactions at one amino acid providing access to highly modified proteins.

Acknowledgements

We thank the Excellence Cluster CiPS^M and SFB749 (Project A4) as well as SFB1032 (Project A5) for financial support. Further support is acknowledged from the Fonds der Chemischen Industrie (predoctoral fellowship to M.S.). We thank Dr. Peter Mayer (LMU Munich) for X-ray analysis.

Keywords: amber suppression • bioorganic chemistry • pyrrolysine • synthetic biology • unnatural amino acid

- [1] C. H. Kim, J. Y. Axup, P. G. Schultz, *Curr. Opin. Chem. Biol.* **2013**, *17*, 412–419.
[2] a) K. Lang, J. W. Chin, *Chem. Rev.* **2014**, *114*, 4764–4806; b) J. M. Harris, R. B. Chess, *Nat. Rev. Drug Discovery* **2003**, *2*, 214–221; c) P. J. Carter, *Exp. Cell Res.* **2011**, *317*, 1261–1269.

- [3] a) B. Hao, W. Gong, T. K. Ferguson, C. M. James, J. A. Krzycki, M. K. Chan, *Science* **2002**, *296*, 1462–1466; b) G. Srinivasan, C. M. James, J. A. Krzycki, *Science* **2002**, *296*, 1459–1462.
[4] a) M. A. Gaston, L. Zhang, K. B. Green-Church, J. A. Krzycki, *Nature* **2011**, *471*, 647–650; b) S. E. Cellitti, W. Ou, H. P. Chiu, J. Grunewald, D. H. Jones, X. Hao, Q. Fan, L. L. Quinn, K. Ng, A. T. Anfora, S. A. Lesley, T. Uno, A. Brock, B. H. Geierstanger, *Nat. Chem. Biol.* **2011**, *7*, 528–530; c) D. G. Longstaff, R. C. Larue, J. E. Faust, A. Mahapatra, L. Zhang, K. B. Green-Church, J. A. Krzycki, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 1021–1026.
[5] a) F. Quittner, A. List, W. Eisenreich, A. Bacher, M. Groll, *Angew. Chem. Int. Ed.* **2012**, *51*, 1339–1342; *Angew. Chem.* **2012**, *124*, 1367–1370; b) F. Quittner, A. List, P. Beck, A. Bacher, M. Groll, *J. Mol. Biol.* **2012**, *424*, 270–282; c) F. Quittner, P. Beck, A. Bacher, M. Groll, *Angew. Chem. Int. Ed.* **2013**, *52*, 7033–7037; *Angew. Chem.* **2013**, *125*, 7171–7175.
[6] a) E. Kaya, K. Gutsmedl, M. Vrabel, M. Muller, P. Thumbs, T. Carell, *Chem-biochem* **2009**, *10*, 2858–2861; b) D. P. Nguyen, H. Lusic, H. Neumann, P. B. Kapadnis, A. Deiters, J. W. Chin, *J. Am. Chem. Soc.* **2009**, *131*, 8720–8721; c) T. Fekner, X. Li, M. M. Lee, M. K. Chan, *Angew. Chem. Int. Ed.* **2009**, *48*, 1633–1635; *Angew. Chem.* **2009**, *121*, 1661–1663; d) E. Kaya, M. Vrabel, C. Deiml, S. Prill, V. S. Fluxa, T. Carell, *Angew. Chem. Int. Ed.* **2012**, *51*, 4466–4469; *Angew. Chem.* **2012**, *124*, 4542–4545; e) M. J. Gattner, M. Vrabel, T. Carell, *Chem. Commun.* **2013**, *49*, 379–381; f) W. Wan, J. M. Tharp, W. R. Liu, *Biochim. Biophys. Acta Proteins Proteomics* **2014**, *1844*, 1059–1070.
[7] a) R. A. Mehl, J. C. Anderson, S. W. Santoro, L. Wang, A. B. Martin, D. S. King, D. M. Horn, P. G. Schultz, *J. Am. Chem. Soc.* **2003**, *125*, 935–939; b) J.-E. Jung, S. Y. Lee, H. Park, H. Cha, W. Ko, K. Sachin, D. W. Kim, D. Y. Chi, H. S. Lee, *Chem. Sci.* **2014**, *5*, 1881–1885.
[8] C. T. Supuran, *Nat. Rev. Drug Discovery* **2008**, *7*, 168–181.
[9] S. Schneider, M. J. Gattner, M. Vrabel, V. Flügel, V. Lopez-Carrillo, S. Prill, T. Carell, *ChemBioChem* **2013**, *14*, 2114–2118.
[10] R. Appel, *Angew. Chem. Int. Ed. Engl.* **1975**, *14*, 801–811; *Angew. Chem.* **1975**, *87*, 863–874.
[11] U. Schöllkopf, U. Groth, C. Deng, *Angew. Chem. Int. Ed. Engl.* **1981**, *20*, 798–799; *Angew. Chem.* **1981**, *93*, 793–795.
[12] P. L. Whitney, *Anal. Biochem.* **1974**, *57*, 467–476.
[13] W. Ou, T. Uno, H. P. Chiu, J. Grunewald, S. E. Cellitti, T. Crossgrove, X. Hao, Q. Fan, L. L. Quinn, P. Patterson, L. Okach, D. H. Jones, S. A. Lesley, A. Brock, B. H. Geierstanger, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 10437–10442.
[14] K. Sivakumar, F. Xie, B. M. Cash, S. Long, H. N. Barnhill, Q. Wang, *Org. Lett.* **2004**, *6*, 4603–4606.

Received: March 11, 2015

Published online on April 2, 2015

7.6 Genetically designed biomolecular capping system for mesoporous silica nanoparticles enables receptor-mediated cell uptake and controlled drug release



Nanoscale

PAPER

[View Article Online](#)
[View Journal](#) | [View Issue](#)

 Cite this: *Nanoscale*, 2016, **8**, 8101

Genetically designed biomolecular capping system for mesoporous silica nanoparticles enables receptor-mediated cell uptake and controlled drug release†

 Stefan Datz,^a Christian Argyo,^a Michael Gattner,^a Veronika Weiss,^a Korbinian Brunner,^a Johanna Bretzler,^a Constantin von Schirnding,^a Adriano A. Torrano,^a Fabio Spada,^a Milan Vrabel,^b Hanna Engelke,^a Christoph Bräuchle,^a Thomas Carell^a and Thomas Bein^{*a}

Effective and controlled drug delivery systems with on-demand release and targeting abilities have received enormous attention for biomedical applications. Here, we describe a novel enzyme-based cap system for mesoporous silica nanoparticles (MSNs) that is directly combined with a targeting ligand *via* bio-orthogonal click chemistry. The capping system is based on the pH-responsive binding of an aryl-sulfonamide-functionalized MSN and the enzyme carbonic anhydrase (CA). An unnatural amino acid (UAA) containing a norbornene moiety was genetically incorporated into CA. This UAA allowed for the site-specific bio-orthogonal attachment of even very sensitive targeting ligands such as folic acid and anandamide. This leads to specific receptor-mediated cell and stem cell uptake. We demonstrate the successful delivery and release of the chemotherapeutic agent Actinomycin D to KB cells. This novel nanocarrier concept provides a promising platform for the development of precisely controllable and highly modular theranostic systems.

Received 18th November 2015,

Accepted 14th February 2016

DOI: 10.1039/c5nr08163g

www.rsc.org/nanoscale

Introduction

The development of effective systems for targeted drug delivery combined with *on demand* release behavior can be considered one of the grand challenges in nanoscience. In particular, porous nanocarriers with high drug loading capacity, immunological stealth behavior and tunable surface properties are promising candidates for biomedical applications such as cancer therapy and bioimaging.^{1–5} Specifically, multifunctional mesoporous silica nanoparticles (MSNs) have great potential in drug delivery applications due to their attractive porosity parameters and the possibility to conjugate release mechanisms for diverse cargos^{6,7} including gold nano-

particles,^{8,9} iron oxide nanocrystals,¹⁰ bio-macromolecules,^{11,12} enzymes,¹³ and polymers.¹⁴ Control over a stimuli-responsive cargo release can be achieved *via* different trigger mechanisms such as redox reactions,¹⁵ pH changes,¹⁶ light-activation,^{6,17} or change in temperature.⁷ Drug delivery vehicles equipped with acid-sensitive capping mechanisms are highly desirable for acidified target environments such as the transition from early to late endosomes, tumors, or inflammatory tissues.

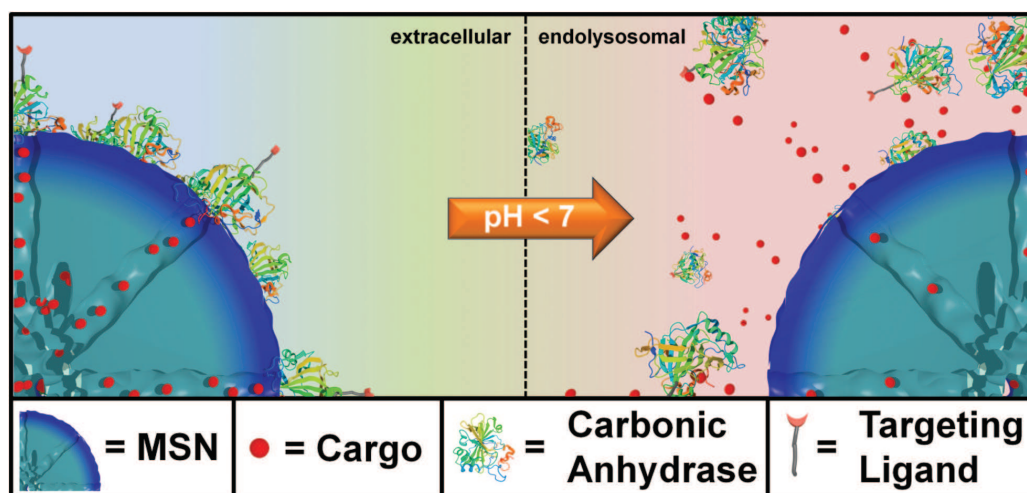
Here, we present genetically designed enzyme-capped MSNs that combine two important prerequisites for advances in drug delivery, namely stimuli-responsive drug release and specific cell targeting (Scheme 1). Specifically, these pH-responsive MSNs consist of a capping structure based on carbonic anhydrase (CA). CA is a model enzyme abundant in humans and animals and generally catalyzes the hydration of carbon dioxide and the dehydration of bicarbonate.¹⁸ It is attached to the silica nanoparticle surface *via* aryl sulfonamide groups. Acting as inhibitors, sulfonamide groups strongly bind to the active site of the CA. This enzyme-sulfonamide binding is reversible depending on the pH, where an acidic medium causes protonation of the sulfonamide, resulting in cleavage of the coordination bond and access to the porous system.¹⁹ The CA gatekeepers were used to exploit the

^aDepartment of Chemistry, Nanosystems Initiative Munich (NIM), Center for Nano Science (CeNS), and Center for Integrated Protein Science Munich (CIPSM), University of Munich (LMU), Butenandtstr. 5-13, 81377 Munich, Germany. E-mail: bein@lmu.de

^bInstitute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Czech Republic

†Electronic supplementary information (ESI) available: Experimental details, further characterization of functionalized MSNs, synthesis of anandamide-tetrazine, results of additional cell studies, dose-dependent MTT assay. See DOI: 10.1039/c5nr08163g





Scheme 1 Schematic illustration of the genetically designed biomolecular pore gating system providing a pH-responsive drug release from mesoporous silica nanoparticles (MSNs). Aryl sulfonamide functionalized MSNs offer pH-dependent reversible attachment of the bulky enzyme carbonic anhydrase, which efficiently blocks the pore entrances to prevent premature cargo release. Furthermore, specific cancer cell targeting can be achieved via site-specific modification of a genetically incorporated norbornene amino acid in the biomolecular gatekeepers.

endosomal pH change as an internal cellular trigger and to gain control over the release of cargo molecules from the mesoporous system.

This stimuli-responsive capping system on MSNs was combined with cell targeting specificity via a bio-orthogonal click chemistry approach. Targeting ligands provide specific binding to certain cell membrane receptors allowing for an enhanced and distinctive cellular uptake of such modified nanocarriers. For example, various cell receptors are over-expressed on cancer cells, which can lead to a preferential receptor-mediated endocytosis of modified MSNs. For the attachment of such targeting ligands exclusively to the outer periphery of the enzyme gatekeepers, we exploited a recently developed method that takes advantage of the pyrrolysine *amber suppression* system followed by bio-orthogonal copper-free click chemistry.^{20–22} This system has already been utilized in applications such as optical gene control.²³ To the best of our knowledge, this is the first time the pyrrolysine *amber suppression* system is used in a combination with porous nanocarriers for specific cell recognition and drug delivery. The incorporation of an unnatural amino acid (UAA) containing a norbornene moiety into CA provides a bio-orthogonal reaction pathway by covalently attaching tetrazine-modified targeting ligands.^{24,25} It has recently been demonstrated that norbornene–tetrazine click chemistry is a favorable synthesis strategy over various other methods including thiol–maleimide reaction and amide formation due to extremely mild and bio-compatible reaction conditions and higher selectivity.²⁶ Here, copper-free click chemistry of norbornene-modified human carbonic anhydrase II with targeting ligands was performed to

prepare folate- and anandamide-modified multifunctional mesoporous silica nanocarriers.²⁷ The anandamide is, due to the *cis*-configured double bonds, a particularly sensitive receptor ligand that requires extremely mild coupling conditions. The targeting system based on folate-modified silica nanocarriers was studied on KB cancer cells, which are known to overexpress the folate receptor FR- α .^{6,28} The targeting system based on anandamide-modified particles was tested on neural stem cells and A431 cells. The combination of on-demand release and specific receptor-mediated cell uptake properties within one multifunctional mesoporous silica nanocarrier system, containing biomolecular valves based on carbonic anhydrase, is anticipated to offer promising potential for controlled drug delivery applications including cancer therapy.

Experimental part

Synthesis of thiol-functionalized MSNs (MSN-SH)

A mixture of TEOS (1.92 g, 9.22 mol) and TEA (14.3 g, 95.6 mmol) was heated to 90 °C for 20 min under static conditions in a polypropylene reactor. Then, a preheated (60 °C) mixture of CTAC (2.41 mL, 1.83 mmol, 25% in H₂O) and NH₄F (100 mg, 0.37 mmol) in bidistilled H₂O (21.7 g, 1.21 mol) was added and the resulting reaction mixture was stirred vigorously (700 rpm) for 30 min while cooling down to room temperature. Afterwards, TEOS (18.2 mg, 92 μ mol) and MPES (18.1 mg, 92 μ mol) were premixed briefly before addition to the reaction mixture. The final reaction mixture was stirred over night at room temperature. After dilution with absolute ethanol



Nanoscale

(100 mL), the nanoparticles were collected by centrifugation (19 000 rpm, 43 146 rcf, 20 min) and redispersed in absolute ethanol. Template extraction was performed in an ethanolic solution of MSNs (100 mL) containing NH_4NO_3 (2 g) which was heated at reflux conditions (90 °C oil bath) for 45 min. This was followed by a second extraction step (90 mL absolute ethanol and 10 mL hydrochloric acid (37%)) under reflux conditions for 45 min (the material was washed with absolute ethanol after each extraction step and collected by centrifugation); finally the particles were redispersed in absolute ethanol and stored as colloidal suspension.

Synthesis of sulfonamide-functionalized MSNs (MSN-phSA)

For the covalent attachment of a sulfonamide derivative to the external particle surface, a thiol-reactive linker was synthesized. 6-Maleimidohexanoic acid *N*-hydroxysuccinimide ester (mal-C₆-NHS, 10 mg, 33 μmol) was dissolved in DMF (500 μL , dry) and was added to an ethanolic solution (15 mL) containing 4-(2-aminoethyl)benzene sulfonamide (6.7 mg, 33 μmol). The resulting reaction mixture was stirred for 1 h at room temperature. Afterwards, thiol-functionalized silica nanoparticles (MSN-SH, 100 mg) in absolute ethanol (10 mL) were added and the mixture was stirred over night at room temperature. Subsequently, the particles were collected by centrifugation (19 000 rpm, 41 146 rcf, 20 min), washed twice with absolute ethanol and were finally redispersed in ethanol (15 mL) to obtain a colloidal suspension.

Cargo loading and particle capping

MSNs (MSN-phSA, 1 mg) were immersed in an aqueous solution of fluorescein (1 mL, 1 mM), DAPI (500 μL , 14.3 mM) or Actinomycin D (500 μL [14 v% DMSO], 140 μM) and stirred over night or for 1 h, respectively. After collection by centrifugation (14 000 rpm, 16 837 rcf, 4 min), the loaded particles were redispersed in a HBSS buffer solution (1 mL) containing carbonic anhydrase (1 mg) and the resulting mixture was allowed to react for 1 h at room temperature under static conditions. The particles were thoroughly washed with HBSS buffer (4 times), collected by centrifugation (5000 rpm, 2200 rcf, 4 min, 15 °C), and finally redispersed in HBSS buffered solution.

Click chemistry of norbornene-containing hCA

MSNs (MSN-phSA, 0.5 mg) were immersed in 500 μL HBSS buffer solution and 0.5 mg norbornene-containing hCA was added. In the meantime, 2.5 μg tetrazine *p*-benzylamine (DMSO stock solution, 0.92 mg mL^{-1}) and 0.41 mg NHS-PEG₂₀₀₀-FA were mixed in 100 μL HBSS and stirred overnight in the dark at room temperature. The solutions were mixed afterwards and stirred for two hours, washed several times and redispersed in 1 mL HBSS buffer. Subsequently, 1 μL Atto633mal (DMF stock solution, 0.5 mg mL^{-1}) was added and the mixture was stirred for 1 hour. The particles were thoroughly washed with HBSS buffer (4 times), collected by centrifugation (5000 rpm, 2200 rcf, 4 min, 15 °C), and finally redispersed in HBSS buffered solution.

Synthesis of Knorb

The norbornene containing amino acid Knorb was synthesized as described in ref. 29.

Mutagenesis of pACA_HCA H36amber

Adapted from ref. 30 with permission from The Royal Society of Chemistry. The amber codon (TAG) was introduced into the expression vector pACA_HCA³¹ at position His36 of the human carbonic anhydrase II gene by blunt end site directed mutagenesis using the primers *forward* HCA H36amber and *reverse* HCA H36amber (see Table 1).

Expression of norbornene-containing HCA

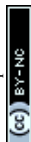
Adapted from ref. 30 with permission from The Royal Society of Chemistry. The expression vector pACA_HCA H36amber was transformed together with pACYC_pylRS Norb, 3xpylT²⁹ which contains the genes of the triple mutant of PylRS and three copies of *pylT* in *E. coli* BL21(DE3) cells (NEB). 1 L of LB medium containing 34 mg L^{-1} chloramphenicol, 100 mg L^{-1} carbenicillin and 2 mM norbornene amino acid Knorb was inoculated with 10 mL of an overnight culture. The cells were stirred at 37 °C until an OD₆₀₀ of 0.9. At this optical density 1 mM ZnSO_4 and 0.1 mM IPTG were added to induce the expression of the HCA H36amber gene. After further 10 h at 37 °C the cells were harvested and stored at −20 °C until further use. The harvested cells were resuspended in washing buffer (25 mM tris; 50 mM Na_2SO_4 ; 50 mM NaClO_4 ; pH 8.8) and disrupted by French Press procedure. The supernatant of the centrifuged lysate was used for sulfonamide affinity protein purification using an ÄKTA purifier system. The self-packed 3 mL column of *p*-aminomethylbenzenesulfonamide-agarose resin (Sigma-Aldrich, A0796) was equilibrated with washing buffer. After binding (0.75 mL min^{-1}) of the protein solution, the column was washed with 7 column volumes of washing buffer. HCA was eluted by lowering the pH by elution buffer (100 mM NaOAc; 200 mM NaClO_4 ; pH 5.6). The protein containing fractions were combined, analyzed by SDS-PAGE, dialyzed against water and lyophilized. Typical yields of the pure norbornene amino acid Knorb containing protein HCA H36Knorb were 20 mg L^{-1} expression medium.

Tryptic digestion and MS/MS of norbornene-containing HCA

Adapted from ref. 30 with permission from The Royal Society of Chemistry. The sequence of HCA II is shown in Table 2.

Table 1 Sequences of the used primers for the generation of expression vector pACA_HCA H36amber. The introduced Amber codon is shown in bold

Name	Sequence
Forward HCA H36amber	5'phosph GTT GAC ATC GAC ACT TAG ACA GCC AAG TAT GAC
Reverse HCA H36amber	5'phosph AGG GGA CTG GCG CTC TCC CTT GG



View Article Online

Nanoscale

Paper

Table 2 Amino acid sequence of HCA II

10	20	30	40	50	60
MAHHWGYGKH	NGPEHWHKDF	PIAKGERQSP	VDIDTHTAKY	DPSLKPLSVS	YDQATSLRIL
70	80	90	100	110	120
NNGHAFNVEF	DDSQDKAVLK	GGPLDGYRL	IQFHFHWGSL	DGQGSEHTVD	KKKYAAELHL
130	140	150	160	170	180
VHWNTKYGDF	CKAVQPPDGL	AVLGIFLKVG	SAKPGLQKV	DVLDSIKTKG	KSADFTNFDP
190	200	210	220	230	240
RGLLPESLDY	WTYPGSLTTP	PLLESVTWIV	LKEPISVSSE	QVLKFRKLN	NGEGEPEELM
250	260				
VDNWRPAQPL	KNRQIKASFK				

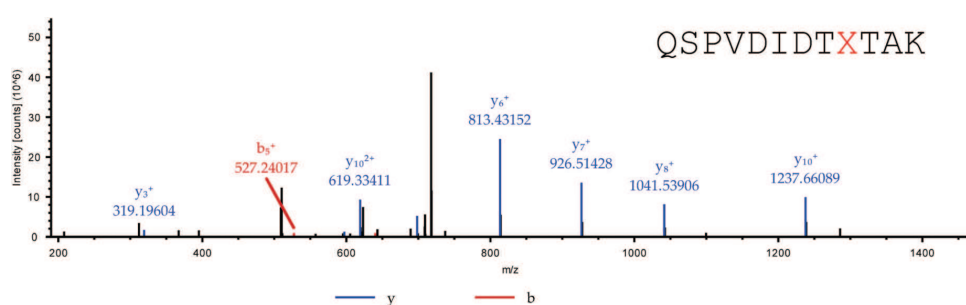


Fig. 1 MS/MS spectrum of the tryptic peptide QSPVDIDTHTAK (X = 4). Parent ion: $[M + 2H]^{2+}_{\text{calc.}} = 726.8829$, $[M + 2H]^{2+}_{\text{obs.}} = 726.8807$ ($\Delta M = 3$ ppm).

Table 3 Expected and identified MS/MS fragments of the tryptic peptide QSPVDIDTHTAK (X = Knorb). Identified fragments are shown in red for b ions and blue for y ions

#1	b^+	Seq.	y^+	#2
1	129.06586	Q		12
2	216.09789	S	1324.69949	11
3	313.15066	P	1237.66746	10
4	412.21908	V	1140.61469	9
5	527.24603	D	1041.54627	8
6	640.33010	I	926.51932	7
7	755.35705	D	813.43525	6
8	856.40473	T	698.40830	5
9	1134.56774	X	597.36062	4
10	1235.61542	T	319.19761	3
11	1306.65254	A	218.14993	2
12		K	147.11281	1

Position His36, which was chosen for the incorporation of amino acid Knorb, is shown in red. The peptide generated after tryptic digestion is emphasized in bold letters. Fig. 1 shows the corresponding MS/MS spectrum. Table 3 shows the expected and identified MS/MS fragments of the relevant tryptic peptide.

Cell culture

HeLa and A431 cells were grown in Dulbecco's modified Eagle's medium (DMEM):F12 (1:1) (Life Technologies) with Glutamax I medium and KB cells in folic acid deficient Roswell

Park Memorial Institute 1640 medium (RPMI 1640, Life Technologies), both supplemented with 10% fetal bovine serum (FBS) at 37 °C in a 5% CO₂ humidified atmosphere. The cells were seeded on collagen A-coated LabTek chambered cover glass (Nunc). For live cell imaging the cells were seeded 24 or 48 h before measuring, at a cell density of 2×10^4 or 1×10^4 cells per cm².

The FGF-2 and EGF dependant neural stem cell line ENC1 was derived from E14 mouse embryonic stem cells and cultured as described.³² ENC1 cells were maintained in gelatine coated flasks and propagated in a 1:1 mixture of Knockout-DMEM (Life Technologies) and Ham's F-12 (Sigma) supplemented with 2 mM GlutaMAX-I (Life Technologies), 100 U ml⁻¹ penicillin (Sigma), 100 µg ml⁻¹ streptomycin (Sigma) 1% N2 and 20 ng ml⁻¹ each of mouse recombinant FGF-2 and EGF (Peprotech). N2 supplement was produced in house as described, with the exception that insulin was of human origin (Sigma I9278) instead of bovine. Stem cells and A431 cells were seeded on ibidi 8-well µ-slides.

In vitro cargo release

Cells were incubated 7–24 h prior to the measurements at 37 °C under a 5% CO₂ humidified atmosphere. Shortly before imaging, the medium was replaced by CO₂-independent medium (Invitrogen). During the measurements all cells were kept on a heated microscope stage at 37 °C. The subsequent imaging was performed as described in the spinning disk confocal microscopy section.



Endosomal compartment staining

To stain the early/late endosome or the lysosome with GFP, commercially available CellLight® staining from lifeTechnologies was used. The cells were simultaneously incubated with MSNs and the BacMam 2.0 reagent. The concentration of the labeling reagent was 25 particles per cell (PCP) of the BacMam 2.0 reagent (*cf.* staining protocol³³). For incubation, the cells stayed at 37 °C under 5% CO₂ humidified atmosphere for 21–24 h before the measurement.

Caspase-3/7 staining

For apoptosis detection commercially available CellEvent™ caspase-3/7 Green Detection Reagent was used. A final concentration of 2.5 μM caspase-3/7 reagent and 0.5 μg mL⁻¹ Hoechst 33342 were added to the cells for 30 min and imaging was performed without further washing steps.

Uptake studies

The functionality of the folic acid targeting ligand was evaluated in a receptor competition experiment. For this purpose, one part of the KB cells was pre-incubated with 3 mM folic acid, to block the receptors, for 2 h at 37 °C under a 5% CO₂ humidified atmosphere. Then the KB cells were incubated with particles for 2/5/8 h at 37 °C under a 5% CO₂ humidified atmosphere. For staining the cell membrane, the cells were incubated with a final concentration of 10 μg mL⁻¹ wheat germ agglutinin Alexa Fluor 488 conjugate for one minute. The cells were washed once with CO₂-independent medium and imaged. For stem cell uptake studies cells were seeded the day prior to incubation. They were incubated for 2 h with free anandamide-tetrazine at a final concentration of 10 μg mL⁻¹. After 2 h 15 μg of particles were added and incubated for another 2 h. Then, the cells were washed 3× with medium containing growth factors and if preincubated free anandamide-tetrazine and incubated until imaging. Immediately before imaging, cell membranes were stained using cell mask deep red (lifetechnologies) and washed with medium.

Spinning disc confocal microscopy

Confocal microscopy for live-cell imaging was performed on a setup based on the Zeiss Cell Observer SD utilizing a Yokogawa spinning disk unit CSU-X1. The system was equipped with a 1.40 NA 100× Plan apochromat oil immersion objective or a 0.45 NA 10× air objective from Zeiss. For all experiments the exposure time was 0.1 s and z-stacks were recorded. DAPI and Hoechst 33342 were imaged with approximately 0.16 W mm⁻² of 405 nm light, GFP and the caspase-3/7 reagent were imaged with approximately 0.48 W mm⁻² of 488 nm excitation light. Atto 633 was excited with 11 mW mm⁻² of 639 nm light. In the excitation path a quad-edge dichroic beamsplitter (FF410/504/582/669-Di01-25 × 36, Semrock) was used. For two color detection of GFP/caspase-3/7 reagent or DAPI/Hoechst 33342 and Atto 633, a dichroic mirror (560 nm, Semrock) and band-pass filters 525/50 and 690/60 (both Semrock) were used in the detection path. Separate images for each fluorescence channel

were acquired using two separate electron multiplier charge coupled device (EMCCD) cameras (PhotometricsEvolve™).

Results and discussion

pH-Responsive MSNs with an average particle size of 150 nm (average pore diameter: 3.8 nm) containing biomolecular valves based on the enzyme carbonic anhydrase (CA, hydrodynamic diameter: 5.5 nm) were synthesized *via* a delayed co-condensation approach.³⁴ An outer functional shell consisting of benzene sulfonamide groups (pHSA) acts as an anchor point for the enzymatic gatekeepers. The formation of the inhibitor-enzyme complex (pHSA-CA) leads to a dense coating at the external particle surface (MSN-pHSA-CA). The characterization of the particle system by dynamic light scattering (DLS), zeta potential measurements, transmission electron microscopy (TEM), nitrogen sorption isotherms, infrared and Raman spectroscopy confirms the successful synthesis of carbonic anhydrase-coated MSNs. *In vial* release experiments demonstrate efficient sealing of the pores with carbonic anhydrase acting as a bulky gatekeeper, preventing premature cargo release and allowing for release upon acid-induced detachment of the capping system (for detailed information about synthesis and characterization see ESI†).

In order to investigate the pH-responsive removal of the bulky gatekeepers from the particles, *in vial* cargo release experiments were performed. We used a custom-made two-compartment system to analyze the time-based release of the fluorescent model cargo fluorescein.¹² After incorporation of fluorescein molecules into the mesoporous system, carbonic anhydrase was added to block the pore entrances. An efficient sealing of the pores and no premature release of the cargo was observed for the sample MSN-pHSA-CA dispersed in HBSS buffer (pH 7.4) at 37 °C (Fig. 2a, closed state, black curve). After 3 h the solution was exchanged and the particles were dispersed in citric-acid phosphate buffer (CAP buffer, pH 5.5). The change to acidic milieu, which simulates the acidification of endosomes, causes a significant increase in fluorescence intensity over time (open state, red curve). Furthermore, we could show the long-term stability of the capping system for more than 16 hours in HBSS buffer and cell medium at pH 7.4 (Fig. 2b). These *in vial* release experiments demonstrate efficient sealing of the pores with carbonic anhydrase acting as a bulky gatekeeper, preventing premature cargo release and allowing for release upon acid-induced detachment of the capping system.

For efficient receptor-mediated cancer cell uptake and selective drug delivery a targeting ligand needs to be implemented. Since the particle surface is covered with bulky enzymes (CA), we aimed for the attachment of the targeting moieties directly to the outer periphery of the enzyme, in order to be accessible for cell receptors. For this approach to be successful, the site of targeting ligand attachment on the enzyme is of key importance. Ideally it should be positioned opposite of the binding site of the enzyme, to prevent blocking of the active site and thus leakage of the capping system. However, site-specific



Paper

View Article Online

Nanoscale

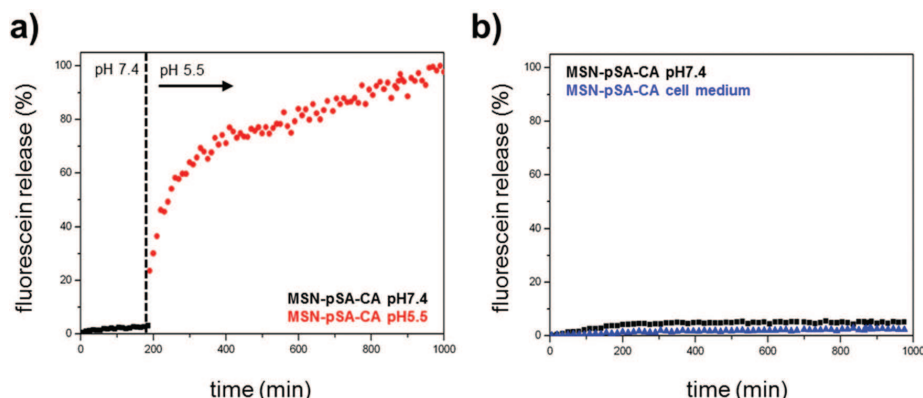


Fig. 2 *In vitro* release kinetics of fluorescein molecules from the enzyme-coated MSNs at different pH values. (a) Sample MSN-pSA-CA features no premature release of the fluorescent cargo molecules in HBSS buffer solution at pH 7.4 (closed state, black curve). After 3 h the medium was changed to slightly acidic milieu (CAP buffer, pH 5.5, red curve) resulting in a significant increase in fluorescence intensity. The gatekeepers are detached from the particle surface upon acidification, causing an efficient and precisely controllable release of fluorescein from the mesoporous system. (b) Long-term stability of the capping system was investigated in HBSS buffer (pH 7.4, black curve) and cell medium (blue curve). No unintended cargo release was observed within about 16 h.

chemical modifications of proteins are highly challenging. Several methods, such as the reaction of thiol groups with maleimide or of lysine chains with activated esters, lack specificity. A more specific method is the incorporation of unnatural amino acids into the protein.^{20,35,36} Among others, the genetic incorporation of UAAs bearing side chains with alkyne,^{37,38} *trans*-cyclooctene,³⁹ cyclooctyne⁴⁰ or norbornene^{20,24} functionalities has been reported previously. Subsequently these residues can be modified specifically and bio-orthogonally, for example by reverse electron-demanding Diels-Alder reactions with tetrazines.^{25,26,39} The natural PylRS/tRNA_{Pyl} pair is perfectly suitable to genetically incorporate UAAs due to its orthogonality to common expression strains. Recently, a norbornene-containing Pyl analogue (Knorb) has been developed by some of us.^{20,21} Here, the synthesis of norbornene-functionalized human carbonic anhydrase II (HCA) was accomplished similar to a previously described procedure yielding HCA H36Knorb.⁴¹ The correct position of the UAA was confirmed by tryptic digestion of the protein followed by HPLC-MS/MS analysis (see ESI†). HCA H36Knorb carrying norbornene on the opposite face of its pHSA-binding site was bound to pHSA-MSN and then treated with an excess of folate-PEG₂₀₀₀-tetrazine (Fig. 3) or anandamide-tetrazine. The excess of the tetrazine reagent could be easily removed by centrifugation of the nanoparticles followed by washing. The efficiency of the folate-targeting system was examined on KB-cells presenting either free or blocked FA-receptors (Fig. 3). For visualization, the cell membrane of the KB cells was stained with WGA488 (green), and the particles were labeled with Atto633 (red). In Fig. 3c–e we present the folic acid receptor blocked cells that were incubated with particles between 2 and 8 h. With increasing incubation time, only a few particles were internalized and unspecific cell uptake was observed only to a

minor degree. In contrast, the cells with available folic acid receptor on their surface (Fig. 3f–h) exhibit a significant and increasing uptake behavior and a considerably higher degree of internalized particles. Thus we could confirm the successful application of bioorthogonal modification of a capping enzyme to act as targeting ligand. We also proved that the genetically modified enzyme capping strategy described here can be used to attach even sensitive ligands like arachidonic acid *via* mild click-chemistry conditions, *e.g.* for the site-specific targeting of neural stem cells and different cancer cells.³² We tested the anandamide-targeting system on neural stem cells and A431 cells. Neural stem cells have anandamide receptors and successfully internalize the anandamide-particles (see ESI, Fig. S9†). A431 cells (epidermoid carcinoma) are also known to over-express the G-protein coupled cannabinoid-based receptor CB2. These receptors can interact with anandamide-functionalized MSNs. Similar to the folate-based targeting experiment, the cannabinoid receptors on the A431 cells were either blocked or free. After 3 h of incubation the receptor-blocked cells internalized just a few anandamide-functionalized MSNs. In contrast, the amount of intracellular particles is clearly much larger in the case of non-blocked cannabinoid receptors (Fig. S10†). The successful experiments with different cell lines and targeting ligands show that the investigated bio-orthogonal attachment concept could be expanded to a variety of enzymes and ligands.

Employing fluorescent live-cell imaging, we investigated the *in vitro* release behavior of encapsulated 4',6-diamidino-2-phenylindole (DAPI) in HeLa cancer cells. The molecular size of DAPI is similar to fluorescein. It was therefore expected to efficiently enter the mesoporous system of the silica nanoparticle. Due to its effective *turn-on* fluorescence upon intercalation into DNA double strands, DAPI is commonly used as



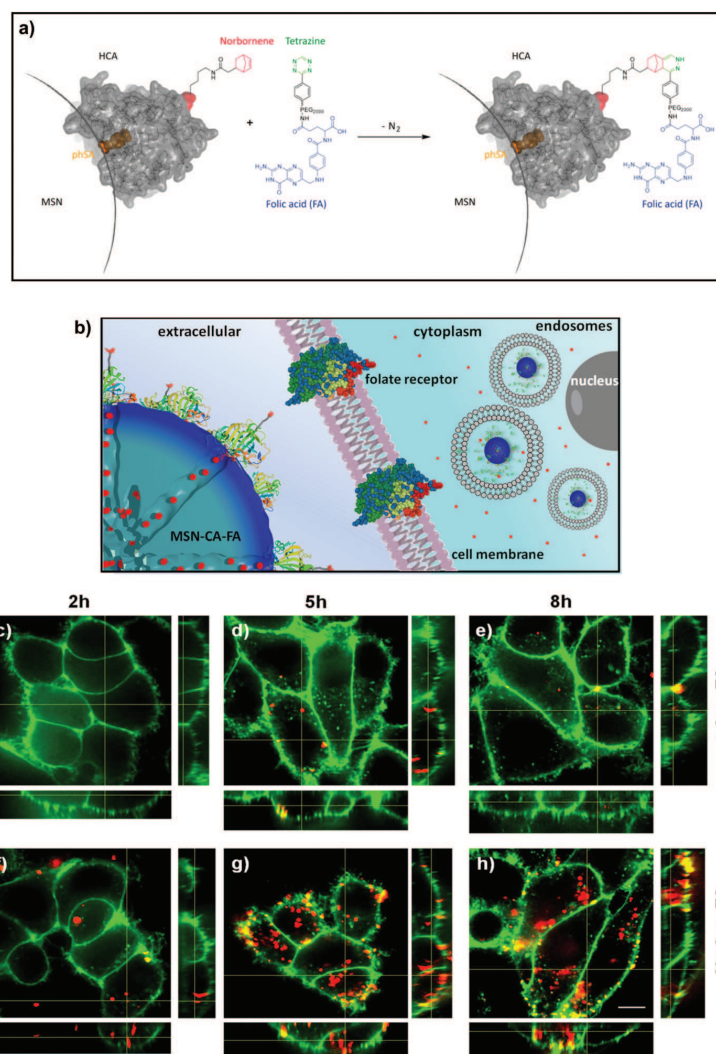


Fig. 3 (a) Norbornene-functionalized carbonic anhydrase (HCA H36Norb) with indicated functionalization site (red) and active site (blue) is able to react in a reversed-electron-demand Diels-Alder reaction with a folate-PEG2000-tetrazine derivative to give HCA-FA. (b) Schematic receptor-mediated uptake of folate-functionalized MSN-CA nanoparticles. (c–e) Nonspecific and (f–h) receptor-mediated endocytosis of MSN-phSA-CA-FA (red) by KB cells (WGA488 membrane staining, green). A specific receptor-mediated cell uptake was observed for MSN-phSA-CA-FA with KB cells (not pre-incubated with FA) after 5 and 8 h incubation at 37 °C (g/h). Incubation of MSN-phSA-CA-FA with FA-pre-incubated KB cells for 2, 5, 8 h at 37 °C showed only minor unspecific cellular uptake over all incubation times (c–e). The scale bar represents 10 μ m.

nuclei counterstain in cell imaging (about 20 fold enhancement in fluorescence intensity).⁴² Since DAPI is cell membrane permeable, free fluorescent dye molecules are able to stain the nucleus within very short time periods (1–5 min), as described in several staining protocols.⁴³ After incorporation of DAPI into the mesoporous system of the silica nanocarriers, the pores were sealed by addition of carbonic anhydrase. The HeLa cells were incubated for a total time period of 24 h with

the loaded particles, which were additionally labeled with Atto 633 dye (red), as depicted in Fig. 4. After 7 h of incubation, MSNs were efficiently taken up by the cells and were found to be located in endosomes. Importantly, almost no staining of the nuclei with DAPI (blue) could be observed at this time point. Only after 15 h, blue fluorescence (even more intensive after 24 h) provided evidence of efficiently released DAPI from the MSNs. Control experiments in which the sample super-



View Article Online

Paper

Nanoscale

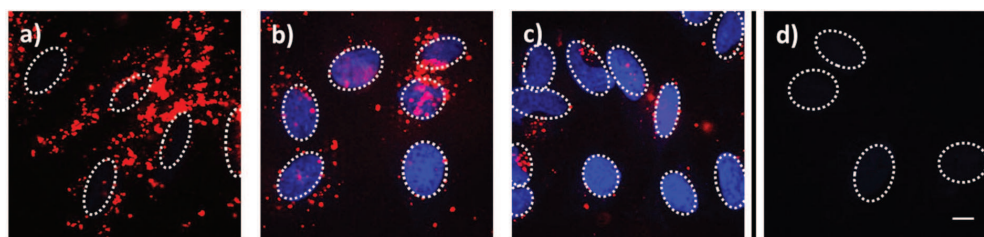


Fig. 4 Fluorescence microscopy of HeLa cells incubated with MSN-phSA-CA nanoparticles loaded with DAPI and labeled with Atto 633 (red) after (a) 7 h, (b) 12 h and (c) 24 h of incubation. The delayed nuclei staining with DAPI (blue) is caused by a time-dependent release of DAPI based on the need for an acidic environment. (d) In a control experiment, the incubation with the supernatant solution (without MSNs) showed no staining of the nuclei with DAPI after 24 h, suggesting that no free DAPI molecules were present in the particle solution. The nuclei are indicated with dashed circles. The scale bar represents 10 μm .

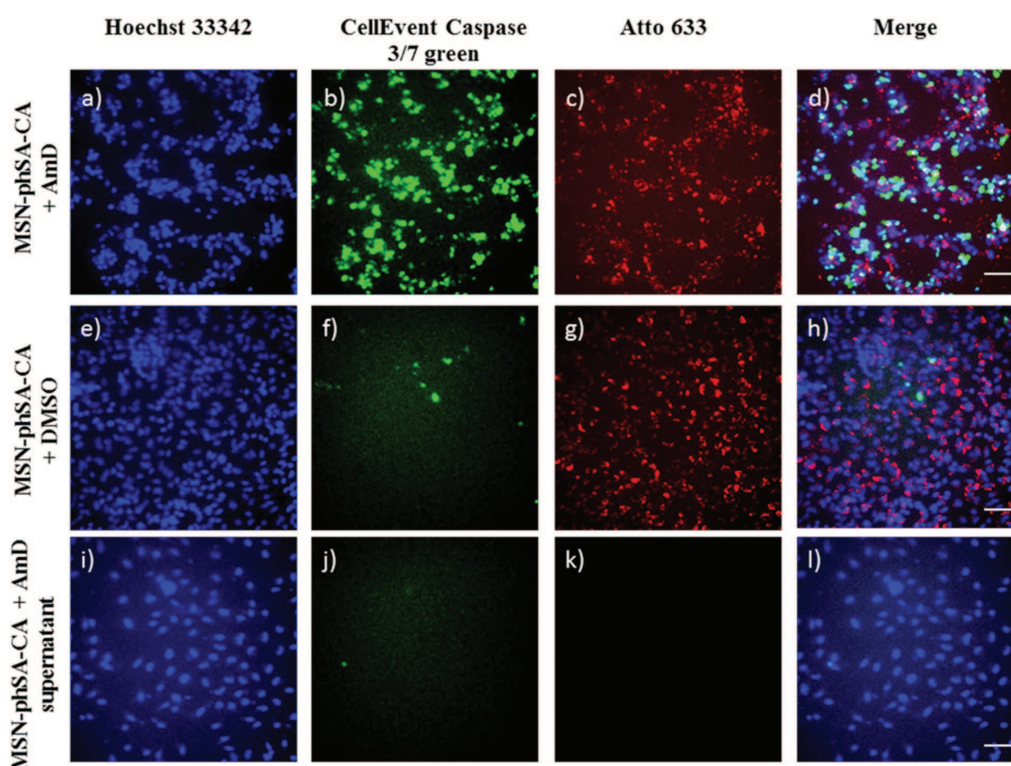


Fig. 5 Representative fluorescence microscopy images of HeLa cells incubated with MSN-phSA-CA nanoparticles loaded with Actinomycin D (AmD; a–d) or DMSO (e–h) and labeled with Atto 633 (red) after 24 h of incubation. As a control, the supernatant of AmD loaded particles after particle separation was incubated with the cells (i–l). Cell nuclei were stained with Hoechst 33342 (blue). For live/dead discrimination CellEvent caspase 3/7 (green) was used. Due to activation of caspase-3/7 in apoptotic cells, DNA can be stained after cleavage of the DNA-binding dye from a binding-inhibiting peptide. MSNs were efficiently taken up by cells (c/d and g/h). Cell death can only be observed for cells treated with AmD loaded MSN-phSA-CA after 24 h of incubation (increased DNA staining in green) (b). In contrast, nanoparticles loaded with DMSO or the sample supernatant do not induce significant apoptosis (almost no DNA-staining) (f and j). The scale bars represent 50 μm .

nanoparticles after particle separation (centrifugation) was added to the HeLa cells showed no significant nuclei staining even after 24 h (Fig. 4d). These cell experiments prove a substantial time-dependent release of DAPI from the mesopores of our

nanocarrier system and also show that no free dye molecules were present in the solution. We suggest that the observed delayed nuclei staining results from a cascaded release mechanism. First, acidification throughout the endosomal pathway



to late endosomes or endolysosomes is of key importance. Only the pH change to mildly acidic values (about 5.5) makes the detachment of the bulky gatekeepers from the MSN hosts possible. Subsequent opening of the pores leads to an efficient cargo release.

Additional co-localization experiments showed the localization of CA-capped nanoparticles in acidic cell compartments after endocytosis (Fig. S8†). To examine the ability of our newly developed MSN drug delivery system to transport chemotherapeutics and to affect cells with their cargo, we incorporated Actinomycin D (AmD), a cytostatic antibiotic, dissolved in DMSO. Free AmD is membrane permeable and induces an uncontrolled cell death within a few hours. MSN-phSA-CA provided intracellular AmD release and caused efficient cell death after 24 h. The delayed reaction demonstrates that AmD was delivered in a controlled manner *via* the particles and released only after acidification of the endosome and subsequent de-capping of the gate-keeper CA. In Fig. 5 cell death is visualized by a caspase 3/7 stain – a marker for apoptotic/dead cells. Control particles loaded with pure DMSO did not induce significant cell death at all, nor did the supernatant solution after particle separation *via* centrifugation (Fig. 5i–l). The results are in good accordance with dose-dependent cell viability studies (Fig. S11†) where the AmD-loaded particles killed HeLa cells effectively after 24 h of incubation ($IC_{50,rel} = 8.3 \mu g mL^{-1}$). This experiment shows the great potential of the MSN-phSA-CA system to efficiently deliver chemotherapeutics to cancer cells. The pH-responsive genetically modified capping system provides the ability to act as a general platform for different targeting ligands and cargos.

Conclusions

We conclude that the novel capping system concept based on pH-responsive detachment of carbonic anhydrase combined with folic acid as targeting ligand allows for highly controllable drug release from porous nanocarriers. Our drug delivery system provides an on-demand release mechanism shown by *in vial* and *in vitro* cargo release experiments. The multifunctional MSNs were efficiently endocytosed in cancer cells and could be located in acidic cell compartments where they released their cargo. Furthermore, the system has an on-board targeting mechanism as demonstrated in additional *in vitro* experiments. The targeting mechanism is attached at a specific site of the capping enzyme preventing interference with the closure mechanism. Our newly developed pH-responsive gatekeepers with genetically designed targeting functions provide a promising platform for the design of versatile and modular drug delivery systems.

Acknowledgements

The authors acknowledge financial support from the Deutsche Forschungsgemeinschaft (DFG) in the context of SFB 749 and SFB 1032, the Excellence Clusters Nanosystems Initiative

Munich (NIM) and Center for Integrated Protein Science Munich (CIPSM), and the Center for Nano Science (CeNS). S. D. thanks the “Dr Klaus-Römer-Stiftung” for financial support. We thank Dr Bastian Rühle for 3D graphics design.

References

- 1 Z. X. Li, J. C. Barnes, A. Bosoy, J. F. Stoddart and J. I. Zink, *Chem. Soc. Rev.*, 2012, **41**, 2590–2605.
- 2 J. L. Vivero-Escoto, I. I. Slowing, B. G. Trewyn and V. S. Y. Lin, *Small*, 2010, **6**, 1952–1967.
- 3 J. M. Rosenholm, V. Mamaeva, C. Sahlgren and M. Linden, *Nanomedicine*, 2012, **7**, 111–120.
- 4 C. Argyo, V. Weiss, C. Bräuchle and T. Bein, *Chem. Mater.*, 2014, **26**, 435–451.
- 5 A. Ott, X. Yu, R. Hartmann, J. Rejman, A. Schutz, M. Ochs, W. J. Parak and S. Carregal-Romero, *Chem. Mater.*, 2015, **27**, 1929–1942.
- 6 S. A. Mackowiak, A. Schmidt, V. Weiss, C. Argyo, C. von Schirnding, T. Bein and C. Bräuchle, *Nano Lett.*, 2013, **13**, 2576–2583.
- 7 A. Schlossbauer, S. Warncke, P. M. E. Gramlich, J. Kecht, A. Manetto, T. Carell and T. Bein, *Angew. Chem., Int. Ed.*, 2010, **49**, 4734–4737.
- 8 E. Aznar, L. Mondragon, J. V. Ros-Lis, F. Sancenon, M. Dolores Marcos, R. Martinez-Manez, J. Soto, E. Perez-Paya and P. Amoros, *Angew. Chem., Int. Ed.*, 2011, **50**, 11172–11175.
- 9 C.-Y. Lai, B. G. Trewyn, D. M. Jeftinija, K. Jeftinija, S. Xu, S. Jeftinija and V. S. Y. Lin, *J. Am. Chem. Soc.*, 2003, **125**, 4451–4459.
- 10 S. Giri, B. G. Trewyn, M. P. Stellmaker and V. S. Y. Lin, *Angew. Chem., Int. Ed.*, 2005, **44**, 5038–5044.
- 11 Z. Luo, K. Cai, Y. Hu, L. Zhao, P. Liu, L. Duan and W. Yang, *Angew. Chem., Int. Ed.*, 2011, **50**, 640–643.
- 12 A. Schlossbauer, J. Kecht and T. Bein, *Angew. Chem., Int. Ed.*, 2009, **48**, 3092–3095.
- 13 W. Zhao, H. Zhang, Q. He, Y. Li, J. Gu, L. Li, H. Li and J. Shi, *Chem. Commun.*, 2011, **47**, 9459–9461.
- 14 R. Liu, X. Zhao, T. Wu and P. Feng, *J. Am. Chem. Soc.*, 2008, **130**, 14418–14419.
- 15 C. Wang, Z. Li, D. Cao, Y.-L. Zhao, J. W. Gaines, O. A. Bozdemir, M. W. Ambrogio, M. Frascioni, Y. Y. Botros, J. I. Zink and J. F. Stoddart, *Angew. Chem., Int. Ed.*, 2012, **51**, 5460–5465.
- 16 Q. Gan, X. Lu, Y. Yuan, J. Qian, H. Zhou, X. Lu, J. Shi and C. Liu, *Biomaterials*, 2011, **32**, 1932–1942.
- 17 S. Carregal-Romero, M. Ochs, P. Rivera-Gil, C. Ganas, A. M. Pavlov, G. B. Sukhorukov and W. J. Parak, *J. Controlled Release*, 2012, **159**, 120–127.
- 18 T. H. Maren, *Phys. Rev.*, 1967, **47**, 595–781.
- 19 P. W. Taylor, R. W. King and A. S. V. Burgen, *Biochemistry*, 1970, **9**, 3894–3902.
- 20 E. Kaya, M. Vrabel, C. Deiml, S. Prill, V. S. Fluxa and T. Carell, *Angew. Chem., Int. Ed.*, 2012, **51**, 4466–4469.



View Article Online

Nanoscale

Paper

- 21 S. Schneider, M. J. Gattner, M. Vrabel, V. Flügel, V. López-Carrillo, S. Prill and T. Carell, *ChemBioChem*, 2013, **14**, 2114–2118.
- 22 L. Davis and J. W. Chin, *Nat. Rev. Mol. Cell Biol.*, 2012, **13**, 168–182.
- 23 J. Hemphill, E. K. Borchardt, K. Brown, A. Asokan and A. Deiters, *J. Am. Chem. Soc.*, 2015, **137**, 5642–5645.
- 24 K. Lang, L. Davis, J. Torres-Kolbus, C. J. Chou, A. Deiters and J. W. Chin, *Nat. Chem.*, 2012, **4**, 298–304.
- 25 T. Plass, S. Milles, C. Koehler, J. Szymanski, R. Müller, M. Wiessler, C. Schultz and E. A. Lemke, *Angew. Chem., Int. Ed.*, 2012, **51**, 4166–4170.
- 26 M. Vrabel, P. Kölle, K. M. Brunner, M. J. Gattner, V. López-Carrillo, R. de Vivie-Riedle and T. Carell, *Chem. – Eur. J.*, 2013, **19**, 13309–13312.
- 27 J. Willibald, J. Harder, K. Sparrer, K.-K. Conzelmann and T. Carell, *J. Am. Chem. Soc.*, 2012, **134**, 12330–12333.
- 28 J. M. Rosenholm, A. Meinander, E. Peuhu, R. Niemi, J. E. Eriksson, C. Sahlgren and M. Linden, *ACS Nano*, 2009, **3**, 197–206.
- 29 E. Kaya, M. Vrabel, C. Deiml, S. Prill, V. S. Fluxa and T. Carell, *Angew. Chem., Int. Ed.*, 2012, **51**, 4466–4469.
- 30 M. J. Gattner, M. Ehrlich and M. Vrabel, *Chem. Commun.*, 2014, **50**, 12568–12571.
- 31 S. K. Nair, T. L. Calderone, D. W. Christianson and C. A. Fierke, *J. Biol. Chem.*, 1991, **266**, 17320–17325.
- 32 K. Brunner, J. Harder, T. Halbach, J. Willibald, F. Spada, F. Gnerlich, K. Sparrer, A. Beil, L. Möckl, C. Bräuchle, K.-K. Conzelmann and T. Carell, *Angew. Chem., Int. Ed.*, 2015, **54**, 1946–1949.
- 33 <http://tools.lifetechnologies.com/content/sfs/manuals/mp10582.pdf>.
- 34 V. Cauda, A. Schlossbauer, J. Kecht, A. Zürner and T. Bein, *J. Am. Chem. Soc.*, 2009, **131**, 11361–11370.
- 35 K. Lang and J. W. Chin, *Chem. Rev.*, 2014, **114**, 4764–4806.
- 36 C. C. Liu and P. G. Schultz, *Annu. Rev. Biochem.*, 2010, **79**, 413–444.
- 37 T. Fekner, X. Li, M. M. Lee and M. K. Chan, *Angew. Chem., Int. Ed.*, 2009, **48**, 1633–1635.
- 38 E. Kaya, K. Gutsmedl, M. Vrabel, M. Muller, P. Thumbs and T. Carell, *ChemBioChem*, 2009, **10**, 2858–2861.
- 39 K. Lang, L. Davis, S. Wallace, M. Mahesh, D. J. Cox, M. L. Blackman, J. M. Fox and J. W. Chin, *J. Am. Chem. Soc.*, 2012, **134**, 10317–10320.
- 40 T. Plass, S. Milles, C. Koehler, C. Schultz and E. A. Lemke, *Angew. Chem., Int. Ed.*, 2011, **50**, 3878–3881.
- 41 M. J. Gattner, M. Ehrlich and M. Vrabel, *Chem. Commun.*, 2014, **50**, 12568–12571.
- 42 M. L. Barcellona, G. Cardiel and E. Gratton, *Biochem. Biophys. Res. Commun.*, 1990, **170**, 270–280.
- 43 <http://tools.lifetechnologies.com/content/sfs/ma+nuals/mp01306.pdf>.



8 Contributions

Declaration of contribution to "*Synthesis of ϵ -N-propionyl-, ϵ -N-butyryl-, and ϵ -N-crotonyl-lysine containing histone H3 using the pyrrolysine system*" (section 7.1)

The project was initiated by Prof. Thomas Carell, Dr. Milan Vrabel and myself. Prof. Thomas Carell supervised the project and gave the scientific direction. Together with him and Dr. Milan Vrabel, we wrote the manuscript and prepared the figures. I performed the cloning of plasmids pPSG IBA3_trxA, pPSG IBA3_trxA N65amber, pPSG IBA35, pPSG IBA35_h3 K9amber, all described protein expressions, protein purifications and protein analyses by fluorescence microscopy, mass spectrometry and western blot.

Declaration of contribution to "*Norbornenes in inverse electron-demand Diels-Alder reactions*" (section 7.2)

The project was initiated by Prof. Thomas Carell and Dr. Milan Vrabel. Prof. Thomas Carell supervised the project and gave the scientific direction. I contributed to this work by supporting the collection of UV/Vis spectroscopic data for the determination of the presented second order rate constants.

Declaration of contribution to "*Structural insights into incorporation of norbornene amino acids for click modification of proteins*" (section 7.3)

The project was initiated by Prof. Thomas Carell, Dr. Sabine Schneider, Dr. Milan Vrabel and myself. Prof. Thomas Carell supervised the project and gave the scientific direction. Together with him, Dr. Sabine Schneider and Dr. Milan Vrabel, we wrote the manuscript. I prepared figure 1, performed the cloning of plasmid pACA_HCA G131amber, protein expression, purification, click-modification and mass spectrometric analyses of all HCA variants. The work related to PylRS crystallization and structure determination was performed by Dr. Veronika Flügel and Dr. Sabine Schneider.

Declaration of contribution to *"Sulfonyl azide-mediated norbornene aziridination for orthogonal peptide and protein labeling"* (section 7.4)

The project was initiated by Dr. Milan Vrabel who also supervised the project and gave the scientific direction. Together with Dr. Michael Ehrlich and Dr. Milan Vrabel we wrote the manuscript and prepared the figures. I supported the peptide labeling experiments and performed all cloning, protein expressions, protein purifications and protein analyses by mass spectrometry described in section 8 and 9 of the according supplementary information (see section 9.2).

Declaration of contribution to *"Orchestrating the biosynthesis of an unnatural pyrrolysine amino Acid for its direct incorporation into proteins inside living cells"* (section 7.5)

The project was initiated by Prof. Thomas Carell, Dr. Michael Ehrlich and myself. Prof. Thomas Carell supervised the project and gave the scientific direction. I contributed to writing the manuscript and prepared the supplemental figures 2 and 3. I developed the biosynthesis and incorporation strategy of ePyl and performed all necessary cloning steps. I performed initial protein expressions, protein purifications as well as the molecular modeling and ePyl incorporation study shown in the supplementary information (see section 9.3).

Declaration of contribution to *"Genetically designed biomolecular capping system for mesoporous silica nanoparticles enables receptor-mediated cell uptake and controlled drug release"* (section 7.6)

The project was initiated by Prof. Thomas Carell, Prof. Thomas Bein, Dr. Christian Argyo and myself. Prof. Thomas Carell and Prof. Thomas Bein supervised the project and gave the scientific direction. I contributed to writing the manuscript and prepared figures 1 and 3a. Together with Johanna Bretzler I performed the norbornene containing HCA protein expression, purification and protein analytics.

9 Selected supplementary information of the original publications

9.1 Selected supplementary information corresponding to section 7.1

Synthesis of ϵ -*N*-propionyl-, ϵ -*N*-butyryl-, and ϵ -*N*-crotonyl-lysine containing histone H3 using the pyrrolysine system

Pages 8-26 and 30 of the supplementary information of the corresponding original publication are shown since they contain methods and data generated by the author of this thesis.

3 Cloning and mutagenesis

3.1 pETDuet_yfp_{TAG}, pylS,T

The cloning procedure of pETDuet_yfp_{TAG} is described in the supporting information of an earlier publication of our group^[1]. In brief, *M. mazei* *pylS* and *pylT* were inserted into the first multiple cloning site using the restriction sites *Bam*HI/*Not*I and *Not*I/*Afl*II, respectively. The gene encoding for the C-terminally StrepII-tagged eYFP Lys114Amber was inserted into the second multiple cloning site using the restriction sites *Nde*I/*Kpn*I. In contrast to the previously published construct, the first T7-promoter was not replaced by a *Trp*-promoter and three copies of the *pylT* gene were inserted instead of one.

3.2 pACyc_3xpylT, pylS

The *Methanosarcina mazei* *pylS* gene was digested using the restriction enzymes *Apa*LI and *Not*I. The 2.2 kb fragment was inserted into pACYC-Duet1 (*Novagen*) using the same sites. The following custom designed DNA-fragment (*GeneArt*) encoding for the *E. coli* pGLN-promoter and a Shine Dalgarno sequence was inserted into this plasmid using the restriction sites *Cl*al/*Nco*I: 5'-TCATCAATCATCCCCATAATCCTTGTTAGATT ATCAATTTTAAAAAACTAACAGTTGTCAGCCTGTCCCGCTTATAATATCATACGCCGTTATACGTTGTTACGCTTTGAGGAAGCC-3'. The resulting construct pACYC-pGLN was subsequently digested with *Not*I and *Pac*I to insert a custom designed DNA-fragment (*GeneArt*) encoding for three copies of *pylT*, which are flanked by the *E. coli* *proK* promoter and terminator, respectively. Sequence of this fragment is as follows, with the *proK*-promoter and -terminator highlighted in bold letters, and *pylT* being underlined: 5'-

GCGGCCTGCTGACTTTCTCGCCGATCAAAAGGCATTTTGCTATTAAGGGATTGACGAGGGCGTATCT
GCGCAGTAAGATGCGCCCCGCATTGGAAACCTGATCATGTAGATCGAATGGACTCTAAATCCGTTCA
GCCGGGTTAGATTCCCGGGGTTTCCGCCATACATGTTATGGAAACCTGATCATGTAGATCGAATGGA
CTCTAAATCCGTTGAGCCGGGTTAGATTCCCGGGGTTTCCGCCATACATGTTATGGAAACCTGATCAT
GTAGATCGAATGGACTCTAAATCCGTTGAGCCGGGTTAGATTCCCGGGGTTTCCGCCA**AATTCGAAA**
AGCCTGCTCAACGAGCAGGCTTTTTGCCTTAAG-3'.

3.3 pPSG IBA3_trxA N65amber

The modified *E. coli* gene *trxA* was cloned by using the *IBA* Stargate System into pPSG IBA3. The gene was amplified from the vector pBAD202 (*Invitrogen*) by using the following primers:

Forward: **AGCGGCTCTTC** AATG GGA TCT GAT AAA ATT ATT CAT CTG ACT GAT G

Reverse: **AGCGGCTCTTC** TCCC CAG GTT AGC GTC GAG GAA CTC TTT C

The PCR-product was inserted into the entry vector IBA51 following the manual of *IBA*. A second cloning step transferred the gene into the destination vector pPSG IBA3. This vector contains an IPTG inducible T7 promotor and a C-terminal Strep-Tag.

The amber codon at position N65 was inserted by blunt-end mutagenesis, using the 5'-phosphorylated primers:

Forward: CCG GGC ACT GCG CCG AAA TAT G

Reverse: CTA GTG ATC GAT GTT CAG TTT TGC AAC GG

3.4 pPSG IBA35_h3 K9amber

The human synthetic gene h3, coding for the histone H3.3 was delivered by *GeneArt*, codon optimized for expression in *E. coli*. The gene was amplified by using the following primers:

Forward: **AGCGGCTCTTC** AATG GCA CGT ACC AAA CAG ACC GCA CGT AAA AG

Forward (+TAG): **AGCGGCTCTTC** AATG GCA CGT ACC AAA CAG ACC GCA CGT TAG AGC ACC G

Reverse: **AGCGGCTCTTC** TCCC TGC ACG TTC ACC ACG AAT ACG ACG TGC

The PCR-product was inserted into the entry vector IBA51 following the manual of *IBA*. A second cloning step transferred the gene into the destination vector pPSG IBA35. This vector contains an IPTG inducible T7 promotor and a N-terminal His₆-Tag.

The amber codon at position K9 was inserted by using a modified Forward primer for the entry vector generation:

Forward (+TAG): **AGCGGCTCTTC** AATG GCA CGT ACC AAA CAG ACC GCA CGT TAG AGC ACC G

4 Testing the incorporation of Kpr, Kbu and Kcr into YFP

Four different *M. mazei* PylRS synthetase variants were tested for their acceptance of the modified lysine amino acids: - wt

- Y306A V401K
- Y306G Y384F I405R
- Y306A Y384F

4.1 Expression of YFP K144Kpr, K144Kbu and K144Kcr

E. coli strain BL21(DE3) (NEB) (*fhuA2 lacZ::T7 gene1 [lon] ompT gal sulA11 R(mcr-73::miniTn10--Tet^S)2 [dcm] R(zgb-210::Tn10--Tet^S) endA1 Δ(mcrC-mrr)114::IS10*) was used for the production of Kpr, Kbu or Kcr containing YFP. Cells were grown in 50 mL LB-medium containing 5 mM of Kbu or Kcr or 10 mM Kpr (37°C, 220 rpm) and 50 mg/L carbenicillin and 34 mg/L Chloramphenicol. When OD₆₀₀=0.6 was reached, expression of *yfp*_{TAG} was induced by addition of 1 mM IPTG. After further 16 h at 30 °C YFP expression was analyzed by fluorescence microscopy.

4.2 Detection of YFP K144Kpr, K144Kbu and K144Kcr

The YFP fluorescence was analyzed by microscopy on a TCS SPE spectral confocal microscope (Leica) with an inverse stand and the HCX FL APO 63x/140-0.60 oil objective (excitation: 488 nm). As shown in Fig. S4.1 the best YFP-read through results for all three acylated lysine derivatives were obtained with the wt-PylRS. Therefore this variant was used for the incorporation of the amino acids into H3.

Electronic Supplementary Material (ESI) for Chemical Communications
This journal is © The Royal Society of Chemistry 2013

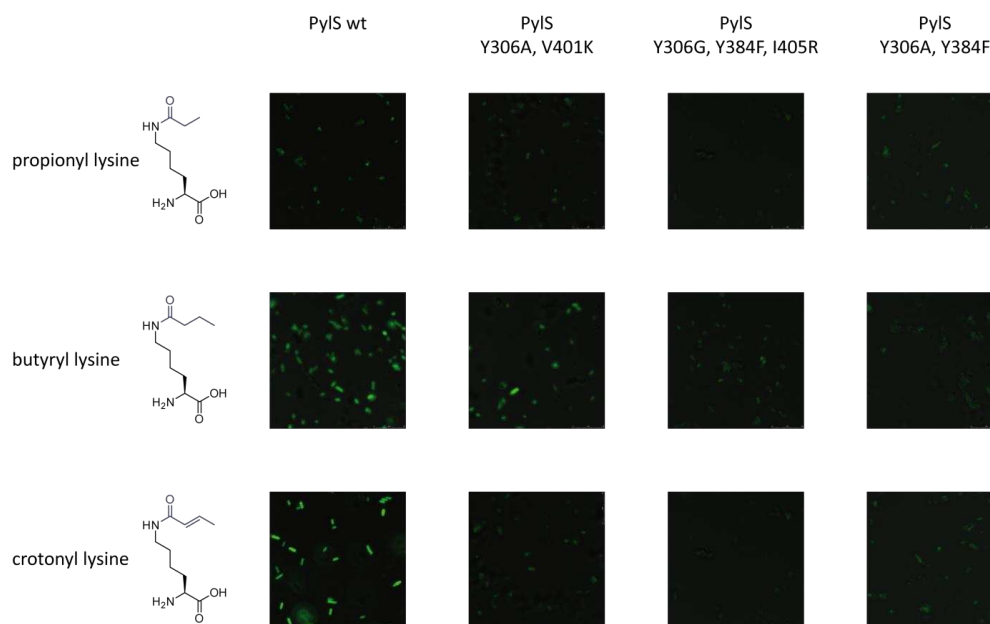


Fig. S4.1. Fluorescence microscopy of *E. coli* transformed with the pETDuet_YFPK114amber reporter system and addition of the three different lysine derivatives Kpr, Kbu and Kcr

5 Incorporation of Kcr into TrxA

5.1 Expression of TrxA N65Kcr

E. coli strain BL21(DE3) were transformed with pACyc_pylTTT, pylS and pPSG IBA3_trxA N65amber. An overnight culture was used for inoculation (to OD₆₀₀ of 0.1) of 2L LB medium containing 5mM Kcr, 50 mg/L carbenicillin and 34 mg/L chloramphenicol. Cells were grown at 37 °C until an OD₆₀₀ of 0.6. Expression of trxA N65Kcr was induced by addition 1 mM IPTG. The cells were shaken for further 16 h at 30 °C. After centrifugation (10.000 x g, 10 min, 4°C) the cells were stored at -20°C or directly purified.

5.2 Purification of TrxA N65Kcr

All purification steps were carried out at 4°C. Cells were resuspended in StrepA buffer (10 mM Tris-HCl pH 8, 1 mM EDTA, 150 mM NaCl) and supplemented with protease inhibitor mix (*Roche*). The cells were lysed using a French press (*Thermo Scientific*). Cell debris was removed by centrifugation (38 000 x g, 30 min, 4 °C) and the supernatant was applied to a 5 mL Strep-Tactin column (*IBA*), equilibrated with StrepA buffer. Proteins were eluted from the column using the same buffer containing 2.5 mM desthiobiotin. The eluted fractions were pooled and concentrated. Typical protein yields of 2 mg per liter expression medium Kcr containing TrxA were achieved. Purified TrxA was equilibrated with StrepA buffer, concentrated and stored at -80°C.

5.3 Intact MS of TrxA N65Kcr

The purified intact protein TrxA N65Kcr was analyzed by HPLC-MS. A self-packed nano C4 column was used for HPLC and mass spectrometry was performed on a LTQ-Orbitrap XL mass spectrometer (*Thermo*). Fig. S5.1 shows the raw mass spectrum. Fig. S5.2 depicts the deconvoluted spectrum. The calculated mass of the protein is 13015.9 Da, the observed mass is 13013.8 Da. Unspecific deacylation of Kcr was not observed at the intact protein level.

Electronic Supplementary Material (ESI) for Chemical Communications
This journal is © The Royal Society of Chemistry 2013

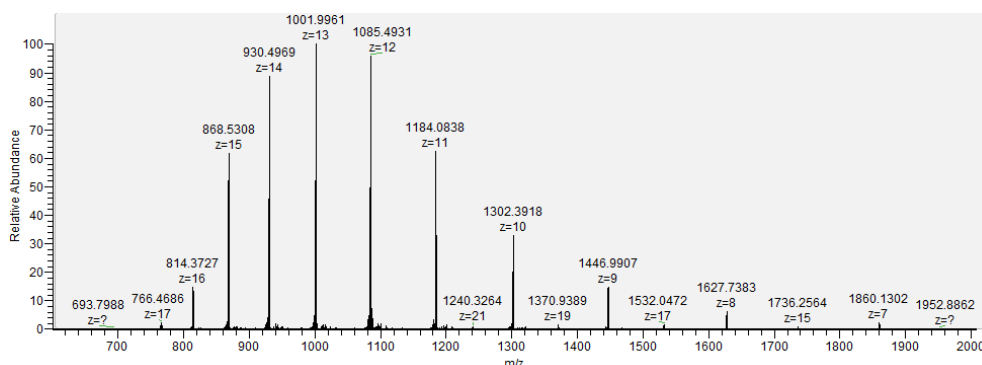


Fig. S5.1. Raw intact mass spectrum of TrxA N65Kcr.

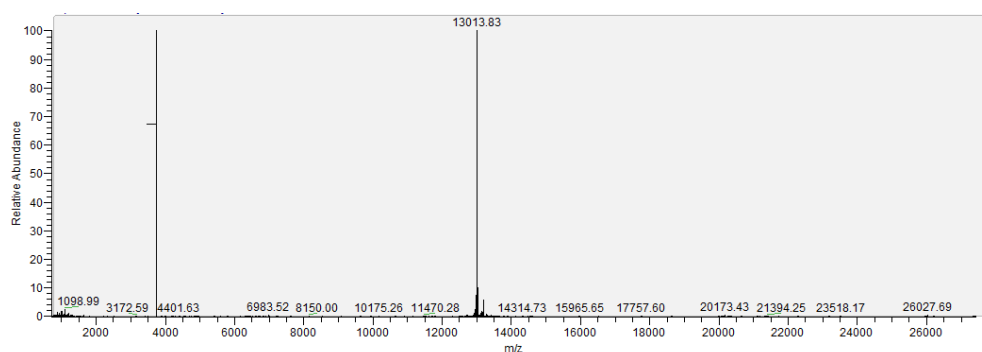


Fig. S5.2. Deconvoluted mass spectrum of TrxA N65Kcr.

5.4 In-Gel chymotryptic digestion of TrxA N65Kcr

The Coomassie Blue stained SDS-gel was washed twice for 10 min with water (bidest.). Protein bands were excised from the gel, cut in small cubes (ca. 1mm³) and transferred to *Eppendorf* tubes. The gel slices were washed by incubation in water (100 µL) for 15 min at RT. Then 50 mM NH₄HCO₃/acetonitrile (each 100 µL) was added and incubated for additional 15 min. All liquids were removed and the gel cubes covered with acetonitrile (100 µL). After removal of acetonitrile the gel particles were rehydrated with 50 mM NH₄HCO₃ (100 µL) for 5 min. This was followed by further addition of 100 µL acetonitrile and incubation for 15 min. The liquids were removed and again 100 µL acetonitrile were added to shrink the particles. After 5 min all acetonitrile was removed. The gel cubes were dried at RT for 10 min. Protein denaturation and reduction were carried out by adding DTT (10 mM in 50 mM NH₄HCO₃, 100 µL) and incubation for 45 min at 56 °C. The tubes were cooled to RT, all liquids were removed and alkylation was performed by addition of 100 µL iodoacetamide (55 mM in 50

mM NH_4HCO_3) and incubation for 30 min at RT (in the dark). Further washing steps (15 min at RT with 50 mM NH_4HCO_3 /acetonitrile, each 100 μL , twice) and addition of acetonitrile (100 μL) resulted in shrinkage of the gel cubes which was completed by removal of all liquids and drying in a vacuum centrifuge (10 min). In-gel digestion was carried out by rehydration of the gel particles in 100 μL 25 mM NH_4HCO_3 containing 0.5 μg chymotrypsin (*Promega*) and incubation at 37 °C overnight. The supernatant was removed and stored separately. For the extraction of the peptides the gel particles were resuspended in 100 μL 25 mM NH_4HCO_3 and sonicated for 15 min before addition of 100 μL acetonitrile and further sonication for 15 min. The supernatant was again removed and stored separately. For quantitative peptide extraction the gel slices were further sonicated with 5% formic acid and acetonitrile (each 100 μL for 15 min). Finally all peptide solutions were combined and concentrated in a vacuum centrifuge. The samples were used for peptide mass analysis after an extended centrifugation step (10 min, 13000 rpm).

5.5 HPLC-MS/MS of chymotryptic digestion of TrxA N65Kcr

For the nano-HPLC ESI mass spectrometry analysis the tryptic digested peptides were loaded onto a *Dionex* C18 Nano Trap Column (100 μm), subsequently eluted and separated by a self packed C18-AQ, 3 μm , 120 Å column. The peptide fragments were analyzed by tandem MS followed by HR-MS using a coupled *Dionex* Ultimate 3000 LTQ-Orbitrap XL MS system (*Thermo*).

The obtained peptide fragment data were analyzed with the SEQUEST algorithm implemented in the software “*Proteome discoverer 1.10.263*” (*Thermo*) against a peptide library of *E. coli* containing the TrxA variant. The search was limited to chymotryptic peptides, two missed cleavage sites, +2 charged monoisotopic precursor ions and a peptide tolerance of <10 ppm. Sequence coverage and probability values of the chymotryptic TrxA N65Kcr digestion sample is shown in Table S5.1.

Table S5.1. Sequence coverage and probability values of the chymotryptic TrxA N65Kcr digestion.

Sample	Sequence coverage	score
TrxA N65Kcr	69.2%	362.3

Table S5.2 sums up the MS/MS evaluation data of the chymotryptic peptide NIDHXPGETAPKY containing either Kcr or K at position X. The amount of deacylation was calculated from the

integrals of the mass signals based on the fact that the ionization properties of the peptides are almost similar.

Table S5.2. MS/MS evaluation data of the chymotryptic peptide NIDHXPGTAPKY of the TrxA N65Kcr digestion.

Sample	$[M+2H]^{2+}$	ΔM	Probability	Ions found
TrxA N65Kcr	704.8649	+ 1.24 ppm	25.78	14/22
TrxA N65K (deacylated)	670.8529	+ 2.06 ppm	32.52	16/22

The MS/MS spectrum of the described $[M+2H]^{2+}$ ions is depicted in Figure S5.3. Sequencing data (MS/MS data) of the relevant peptide fragment NIDHXPGTAPKY of Kcr containing TrxA (X=Kcr, Table S5.3) proof the insertion of the amino acid Kcr at the correct position (N65Kcr). Fig. S5.4 and table S5.4 show the relevant peptide fragments of deacylated Trx N65Kcr (Trx N65K). As fragmentation of the peptide can occur starting either with the N-terminus (b ion series) or the C-terminus (y ion series) both ion series can always be found.

Electronic Supplementary Material (ESI) for Chemical Communications
This journal is © The Royal Society of Chemistry 2013

TrxA N65Kcr

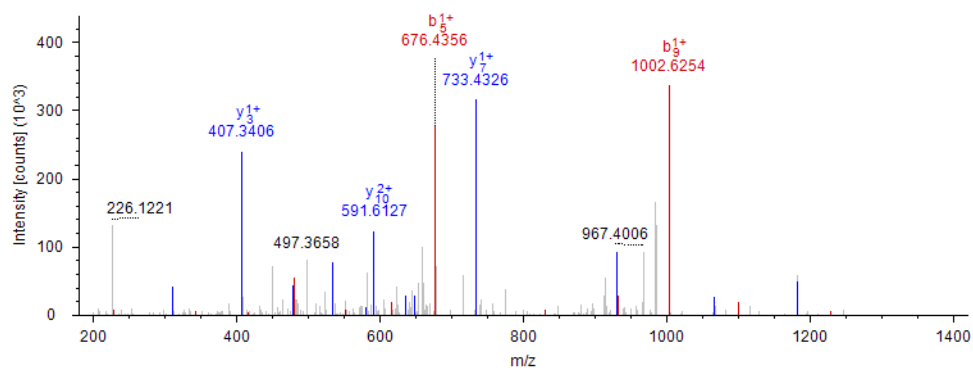


Fig. S5.3. MS/MS spectrum of the chymotryptic peptide NIDHXPGTAPKY (X = Kcr) of the TrxA N65Kcr digestion.

Table S5.3. Assigned MS/MS fragments resulting from fragmentation of the chymotryptic peptide NIDHXPGTAPKY (X = Kcr) of the TrxA N65Kcr digestion.

#1	b(1+)	b(2+)	Seq.	y(1+)	y(2+)	#2
1	115.05021	58.02874	N			12
2	228.13428	114.57078	I	1294.67900	647.84314	11
3	343.16123	172.08425	D	1181.59493	591.30110	10
4	480.22014	240.61371	H	1066.56798	533.78763	9
5	676.34127	338.67427	N-KCr	929.50907	465.25817	8
6	773.39404	387.20066	P	733.38794	367.19761	7
7	830.41551	415.71139	G	636.33517	318.67122	6
8	931.46319	466.23523	T	579.31370	290.16049	5
9	1002.50031	501.75379	A	478.26602	239.63665	4
10	1099.55308	550.28018	P	407.22890	204.11809	3
11	1227.64805	614.32766	K	310.17613	155.59170	2
12			Y	182.08116	91.54422	1

Electronic Supplementary Material (ESI) for Chemical Communications
This journal is © The Royal Society of Chemistry 2013

TrxA N65K (deacylated Kcr)

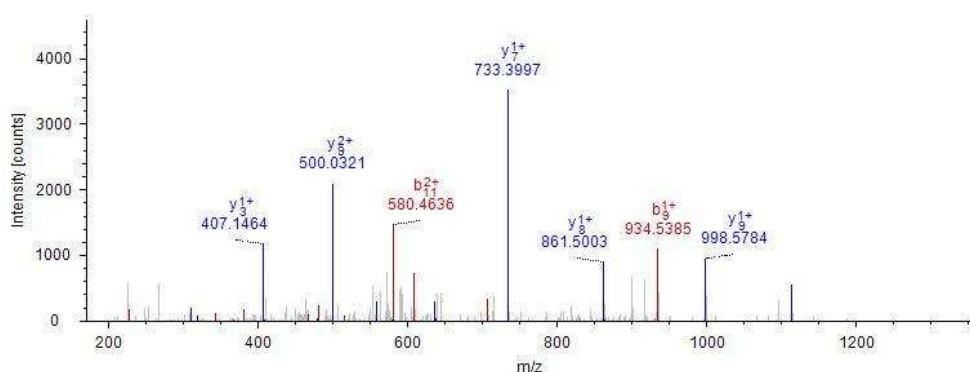


Fig. S5.4. MS/MS spectrum of the deacylated chymotryptic peptide NIDHXPGTAPKY (X = K) of the TrxA N65Kcr digestion.

Table S5.4. Assigned MS/MS fragments resulting from fragmentation of the deacylated chymotryptic peptide NIDHXPGTAPKY (X = K) of the TrxA N65Kcr digestion.

#1	b(1+)	b(2+)	Seq.	y(1+)	y(2+)	#2
1	115.05021	58.02874	N			12
2	228.13428	114.57078	I	1226.65284	613.83006	11
3	343.16123	172.08425	D	1113.56877	557.28802	10
4	480.22014	240.61371	H	998.54182	499.77455	9
5	608.31510	304.66119	N-Asn->Lys	861.48291	431.24509	8
6	705.36787	353.18757	P	733.38794	367.19761	7
7	762.38934	381.69831	G	636.33517	318.67122	6
8	863.43702	432.22215	T	579.31370	290.16049	5
9	934.47414	467.74071	A	478.26602	239.63665	4
10	1031.52691	516.26709	P	407.22890	204.11809	3
11	1159.62188	580.31458	K	310.17613	155.59170	2
12			Y	182.08116	91.54422	1

6 Incorporation of Kpr, Kbu and Kcr into H3

6.1 Expression of H3 wt, K9Kpr, K9Kbu, K9Kcr

E. coli strain BL21(DE3) was transformed with pACyc_pylTTT, pylS and pPSG IBA35_h3 K9amber. An overnight culture was used for inoculation (to OD₆₀₀ of 0.1) of 0.5 L LB medium containing 5 mM Kcr or Kbu or 10 mM Kpr, 50 mg/L carbenicillin and 34 mg/L chloramphenicol. Cells were grown at 37 °C until an OD₆₀₀ of 0.6. Expression of h3 K9Kmod was induced by addition 1 mM IPTG. Optionally 20 mM nicotinamide (NAM) were added at this point to inhibit a possible deacylation of the incorporated modified amino acids. The cells were shaken for further 16 h at 30 °C. After centrifugation (10.000 x g, 10 min, 4°C) the cells were stored at -20°C or directly purified.

6.2 Purification of H3 wt, K9Kpr, K9Kbu, K9Kcr

All purification steps were carried out at room temperature. Cells were resuspended in HisA buffer (50 mM Tris-HCl, 500 mM NaCl, 20 mM imidazol, pH 7.5) supplemented with 7 M urea. The cells were lysed using a French press (*Thermo Scientific*). Cell debris was removed by centrifugation (38 000 x g, 30 min, rt) and the supernatant was applied to a 1 mL HisTrap FF column (*GE*), equilibrated with HisA buffer. Proteins were eluted from the column using the same buffer containing 250 mM imidazol. The eluted fractions were pooled and concentrated. Typical protein yields of 1 mg per liter expression medium Kcr or Kbu containing H3 were achieved. The yield of Kpr containing H3 was typically lower at 0.5 mg/L expression medium. H3 wt protein yield was at 4-5 mg/L. Purified H3 was optionally dialyzed twice (2 h, than over night) against 2 L StrepA buffer (10 mM Tris-HCl pH 8, 1 mM EDTA, 150 mM NaCl) supplemented with 5 mM DTT at 4 °C. concentrated and stored at -80°C.

6.3 In-Gel Propionylation and butyrylation and tryptic digestion of H3 variants H3 wt, K9Kpr, K9Kbu, K9Kcr

The first washing steps of this procedure were performed in the same way as for the chymotryptic digestion described in 5.5.

Instead of the reduction and alkylation, the capping of unmodified lysine residues was carried out^[2]. After this procedure trypsin cleaves exclusively after arginine residues. Consequently longer peptides are generated from the digestion which can be analyzed.

6 Incorporation of Kpr, Kbu and Kcr into H3

6.1 Expression of H3 wt, K9Kpr, K9Kbu, K9Kcr

E. coli strain BL21(DE3) was transformed with pACyc_pylTTT, pylS and pPSG IBA35_h3 K9amber. An overnight culture was used for inoculation (to OD₆₀₀ of 0.1) of 0.5 L LB medium containing 5 mM Kcr or Kbu or 10 mM Kpr, 50 mg/L carbenicillin and 34 mg/L chloramphenicol. Cells were grown at 37 °C until an OD₆₀₀ of 0.6. Expression of h3 K9Kmod was induced by addition 1 mM IPTG. Optionally 20 mM nicotinamide (NAM) were added at this point to inhibit a possible deacylation of the incorporated modified amino acids. The cells were shaken for further 16 h at 30 °C. After centrifugation (10.000 x g, 10 min, 4°C) the cells were stored at -20°C or directly purified.

6.2 Purification of H3 wt, K9Kpr, K9Kbu, K9Kcr

All purification steps were carried out at room temperature. Cells were resuspended in HisA buffer (50 mM Tris-HCl, 500 mM NaCl, 20 mM imidazol, pH 7.5) supplemented with 7 M urea. The cells were lysed using a French press (*Thermo Scientific*). Cell debris was removed by centrifugation (38 000 x g, 30 min, rt) and the supernatant was applied to a 1 mL HisTrap FF column (*GE*), equilibrated with HisA buffer. Proteins were eluted from the column using the same buffer containing 250 mM imidazol. The eluted fractions were pooled and concentrated. Typical protein yields of 1 mg per liter expression medium Kcr or Kbu containing H3 were achieved. The yield of Kpr containing H3 was typically lower at 0.5 mg/L expression medium. H3 wt protein yield was at 4-5 mg/L. Purified H3 was optionally dialyzed twice (2 h, than over night) against 2 L StrepA buffer (10 mM Tris-HCl pH 8, 1 mM EDTA, 150 mM NaCl) supplemented with 5 mM DTT at 4 °C. concentrated and stored at -80°C.

6.3 In-Gel Propionylation and butyrylation and tryptic digestion of H3 variants H3 wt, K9Kpr, K9Kbu, K9Kcr

The first washing steps of this procedure were performed in the same way as for the chymotryptic digestion described in 5.5.

Instead of the reduction and alkylation, the capping of unmodified lysine residues was carried out^[2]. After this procedure trypsin cleaves exclusively after arginine residues. Consequently longer peptides are generated from the digestion which can be analyzed.

Another advantage is the possible comparison of originally modified with deacylated peptides.

The dried gel pieces were incubated with 20 μL of propionic or butyric anhydride and 40 μL of 2 M ammonium hydroxide for 30 min at 51 $^{\circ}\text{C}$. This step was repeated once. Afterwards the gel pieces were washed as described in 5.5. The tryptic digestion was performed by adding 100 μL of a 5 ng/ μL trypsin (*Promega*) solution in 25 mM ammonium hydrogencarbonate. The digestion solution was incubated over night at 37 $^{\circ}\text{C}$. Afterwards the capping procedure was repeated twice as described after the digestion to also cap the N-termini generated during digestion.

The peptides were extracted from the gel pieces and analyzed by HPLC-MS/MS in the same way as described for the chymotryptic peptides in 5.5.

6.4 HPLC-MS/MS of propionylated or butyrylated tryptic peptides of H3 variants H3 wt, K9Kpr, K9Kbu, K9Kcr

For the nano-HPLC ESI mass spectrometry analysis the tryptic digested peptides were loaded onto a *Dionex* C18 Nano Trap Column (100 μm), subsequently eluted and separated by a self packed C18-AQ, 3 μm , 120 \AA column. The peptide fragments were analyzed by HR-MS followed by tandem MS using a coupled *Dionex* Ultimate 3000 LTQ-Orbitrap XL MS system (*Thermo*).

The obtained peptide fragment data were analyzed with the SEQUEST algorithm implemented in the software "*Proteome discoverer 1.1.0.263*" (*Thermo*) against a peptide library of *E. coli* containing the protein sequence of human histone H3.3. The search was limited to tryptic peptides, two missed cleavage sites, +2 charged monoisotopic precursor ions and a peptide tolerance of <10 ppm. The sequence coverage and probability values of the capped tryptic H3 wt, H3 K9Kpr, Kbu and Kcr digestion samples are shown in Table S6.1.

Table S6.1. Sequence coverage and probability values of the capped tryptic H3 K9Kpr, Kbu and Kcr digestion samples.

Sample	capping	Sequence coverage	score
H3 wt	propionylation	62.05%	73.51
H3 wt	butyrylation	61.03%	52.72
H3 K9Kpr	butyrylation	54.41%	105.58
H3 K9Kbu	propionylation	60.29%	47.37
H3 K9Kcr	propionylation	48.53%	21.44

Table S6.2 sums up the MS/MS evaluation data of the tryptic peptide XSTGGKAPR containing either K (capped), Kpr, Kbu or Kcr at position X. The amount of deacylation mentioned in the main paper was calculated from the integrals of the mass signals based on the fact that the ionization properties of the peptides are almost similar.

Table S6.2. MS/MS evaluation data of the tryptic peptide XSTGGKAPR (X = K, Kpr, Kbu or Kcr) of the capped H3 digestion.

Sample	capping	[M+2H] ²⁺	ΔM	Probability	Ions found
H3 wt	propionylation	535.3057	3.86 ppm	45.19	14/16
H3 wt	butyrylation	556.3284	2.09 ppm	24.71	14/16
H3 K9Kpr	butyrylation	549.3199	1.07 ppm	10.75	13/16
H3 K9Kbu	propionylation	542.3140	4.65 ppm	17.23	13/16
H3 K9Kcr	propionylation	541.3046	1.79 ppm	21.44	12/16

MS Spectra of the described [M+2H]²⁺ ions are depicted in Figures S6.1-S6.5. Sequencing data (MS/MS data) of the relevant peptide fragment XSTGGKAPR of modified lysine containing H3 (X = Kpr, Kbu or Kcr, Table S6.3-S6.7) proof the insertion of the modified amino acids at the correct position in H3 (K9). As fragmentation of the peptide can occur starting either with the N-terminus (b ion series) or the C-terminus (y ion series) both ion series can always be found.

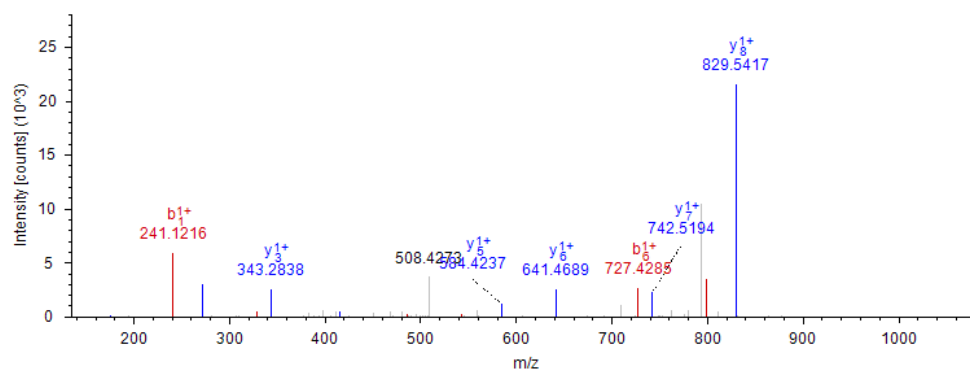
H3 wt propionylated

Fig. S6.1. MS/MS spectrum of the propionyl capped tryptic peptide XSTGGKAPR (X = K, capped by propionylation) of the H3 wt digestion.

Table S6.3. Assigned MS/MS fragments resulting from fragmentation of the propionyl capped tryptic peptide XSTGGKAPR (X = K, capped by propionylation) of the H3 wt digestion.

#1	b(1+)	b(2+)	Seq.	y(1+)	y(2+)	#2
1	241.15465	121.08096	K-Propionyl N-term-Propionyl			9
2	328.18668	164.59698	S	829.45267	415.22997	8
3	429.23436	215.12082	T	742.42064	371.71396	7
4	486.25583	243.63155	G	641.37296	321.19012	6
5	543.27730	272.14229	G	584.35149	292.67938	5
6	727.39847	364.20287	K-Propionyl	527.33002	264.16865	4
7	798.43559	399.72143	A	343.20885	172.10806	3
8	895.48836	448.24782	P	272.17173	136.58950	2
9			R	175.11896	88.06312	1

Electronic Supplementary Material (ESI) for Chemical Communications
This journal is © The Royal Society of Chemistry 2013

H3 wt butyrylated

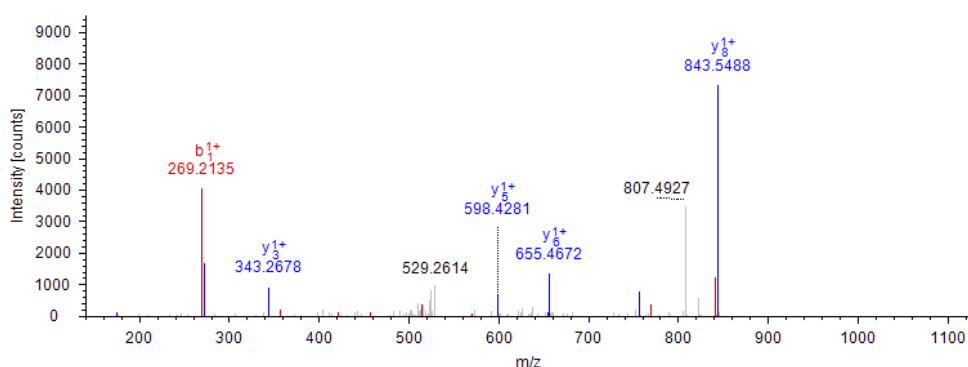


Fig. S6.2. MS/MS spectrum of the butyryl capped tryptic peptide XSTGGKAPR (X = K, capped by butyrylation) of the H3 wt digestion.

Table S6.4. Assigned MS/MS fragments resulting from fragmentation of the butyryl capped tryptic peptide XSTGGKAPR (X = K, capped by butyrylation) of the H3 wt digestion.

#1	b(1+)	b(2+)	Seq.	y(1+)	y(2+)	#2
1	269.18605	135.09666	K-Butyryl N-term-Butyryl			9
2	356.21808	178.61268	S	843.46837	422.23782	8
3	457.26576	229.13652	T	756.43634	378.72181	7
4	514.28723	257.64725	G	655.38866	328.19797	6
5	571.30870	286.15799	G	598.36719	299.68723	5
6	769.44557	385.22642	K-Butyryl	541.34572	271.17650	4
7	840.48269	420.74498	A	343.20885	172.10806	3
8	937.53546	469.27137	P	272.17173	136.58950	2
9			R	175.11896	88.06312	1

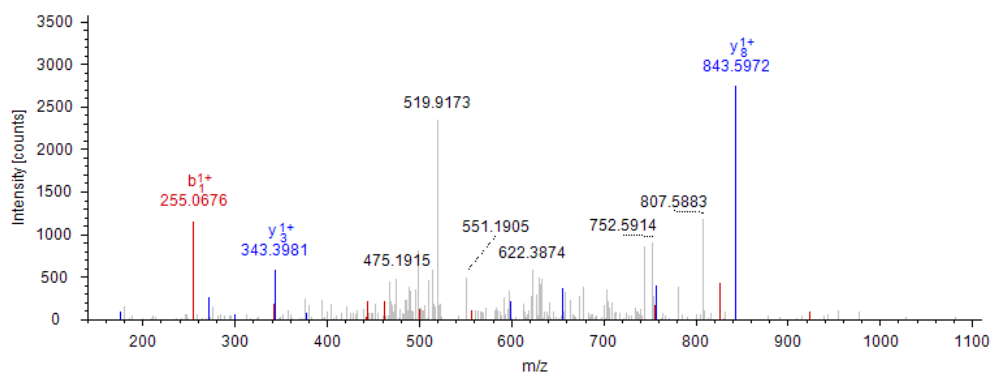
H3 K9Kpr butyrylated

Fig. S6.3. MS/MS spectrum of the butyryl capped tryptic peptide XSTGGKAPR (X = Kpr) of the H3 K9Kpr digestion.

Table S6.5. Assigned MS/MS fragments resulting from fragmentation of the butyryl capped tryptic peptide XSTGGKAPR (X = Kpr) of the H3 K9Kpr digestion.

#1	b(1+)	b(2+)	Seq.	y(1+)	y(2+)	#2
1	255.17035	128.08881	K-Butyryl N-term-Propionyl			9
2	342.20238	171.60483	S	843.46837	422.23782	8
3	443.25006	222.12867	T	756.43634	378.72181	7
4	500.27153	250.63940	G	655.38866	328.19797	6
5	557.29300	279.15014	G	598.36719	299.68723	5
6	755.42987	378.21857	K-Butyryl	541.34572	271.17650	4
7	826.46699	413.73713	A	343.20885	172.10806	3
8	923.51976	462.26352	P	272.17173	136.58950	2
9			R	175.11896	88.06312	1

Electronic Supplementary Material (ESI) for Chemical Communications
This journal is © The Royal Society of Chemistry 2013

H3 K9Kbu propionylated

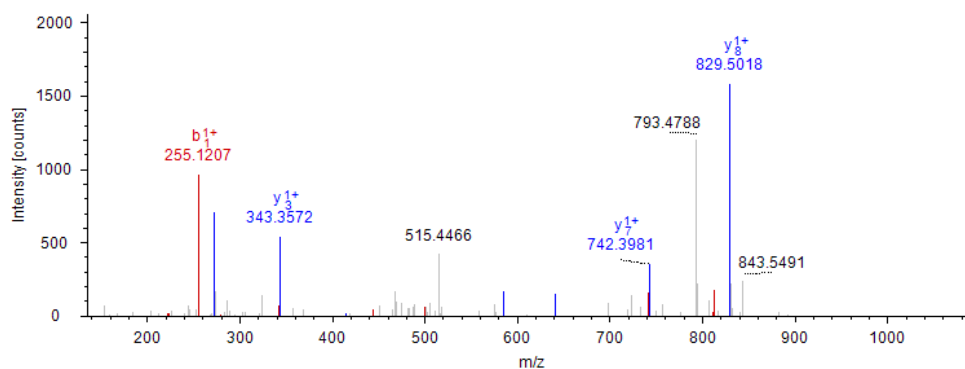


Fig. S6.4. MS/MS spectrum of the propionyl capped tryptic peptide XSTGGKAPR (X = Kbu) of the H3 K9Kbu digestion.

Table S6.6. Assigned MS/MS fragments resulting from fragmentation of the propionyl capped tryptic peptide XSTGGKAPR (X = Kbu) of the H3 K9Kbu digestion.

#1	b(1+)	b(2+)	Seq.	y(1+)	y(2+)	#2
1	255.17035	128.08881	K- Butyryl N-term- Propionyl			9
2	342.20238	171.60483	S	829.45267	415.22997	8
3	443.25006	222.12867	T	742.42064	371.71396	7
4	500.27153	250.63940	G	641.37296	321.19012	6
5	557.29300	279.15014	G	584.35149	292.67938	5
6	741.41417	371.21072	K-Propionyl	527.33002	264.16865	4
7	812.45129	406.72928	A	343.20885	172.10806	3
8	909.50406	455.25567	P	272.17173	136.58950	2
9			R	175.11896	88.06312	1

Electronic Supplementary Material (ESI) for Chemical Communications
This journal is © The Royal Society of Chemistry 2013

H3 K9Kcr propionylated

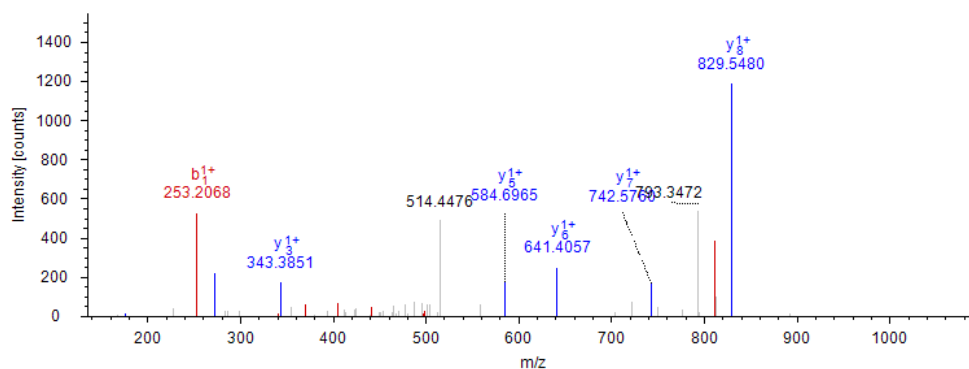


Fig. S6.5. MS/MS spectrum of the propionyl capped tryptic peptide XSTGGKAPR (X = Kcr) of the H3 K9Kcr digestion.

Table S6.7. Assigned MS/MS fragments resulting from fragmentation of the propionyl capped tryptic peptide XSTGGKAPR (X = Kcr) of the H3 K9Kcr digestion.

#1	b(1+)	b(2+)	Seq.	y(1+)	y(2+)	#2
1	253.15465	127.08096	K-Crotonyl N-term-Propionyl			9
2	340.18668	170.59698	S	829.45267	415.22997	8
3	441.23436	221.12082	T	742.42064	371.71396	7
4	498.25583	249.63155	G	641.37296	321.19012	6
5	555.27730	278.14229	G	584.35149	292.67938	5
6	739.39847	370.20287	K-Propionyl	527.33002	264.16865	4
7	810.43559	405.72143	A	343.20885	172.10806	3
8	907.48836	454.24782	P	272.17173	136.58950	2
9			R	175.11896	88.06312	1

Electronic Supplementary Material (ESI) for Chemical Communications
This journal is © The Royal Society of Chemistry 2013

7 Western Blot

The antibodies against Kpr, Kbu and Kcr were purchased from *PTM biolabs*. The western blot was performed following the procedures recommended by *PTM biolabs*. As shown in Figure S7.1 the H3 K9Kpr, H3 K9Kbu and H3 K9Kcr histone modifications can be selectively labeled using the corresponding antibodies when compared to the wild type histone (H3 K9). In a cross experiment every antibody was tested against every histone modification (H3 K9Kpr, H3 K9Kbu and H3 K9Kcr). This experiment revealed that the antibodies also recognize the related modifications (e.g. H3 K9Kpr is recognized also by anti Kbu antibody). This cross labeling activity could be a result of the high concentrations of purified histones used for the western blot. The results are summed up in Fig. S7.2.

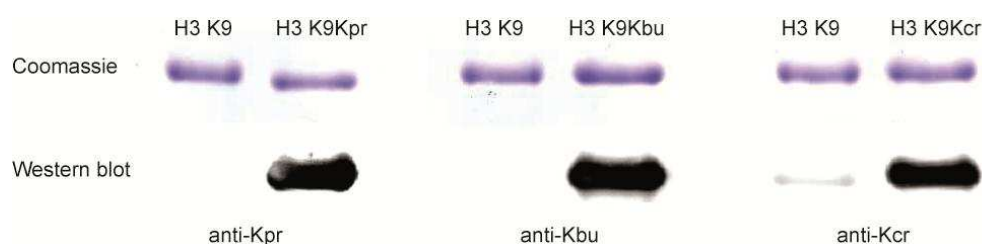


Fig. S7.1. Western Blot of the three different antibodies against different modifications

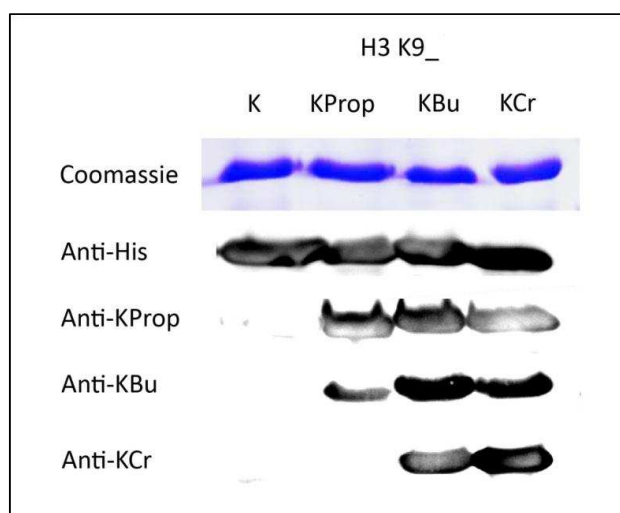


Fig. S7.2. Western Blot of the three different antibodies against different modifications

Electronic Supplementary Material (ESI) for Chemical Communications
This journal is © The Royal Society of Chemistry 2013

9 Literature

- [1] E. Kaya, K. Gutsmedl, M. Vrabel, M. Muller, P. Thumbs, T. Carell, *ChemBioChem* **2009**, *10*, 2858-2861.
- [2] B. A. Garcia, S. Mollah, B. M. Ueberheide, S. A. Busby, T. L. Muratore, J. Shabanowitz, D. F. Hunt, *Nat. Protoc.* **2007**, *2*, 933-938.

9.2 Selected supplementary information corresponding to section 7.4

Sulfonyl azide-mediated norbornene aziridination for orthogonal peptide and protein labeling

Pages S11-S13 and S17-S22 of the supplementary information of the corresponding original publication are shown since they contain methods and data generated by the author of this thesis.

5. Peptide labeling experiments

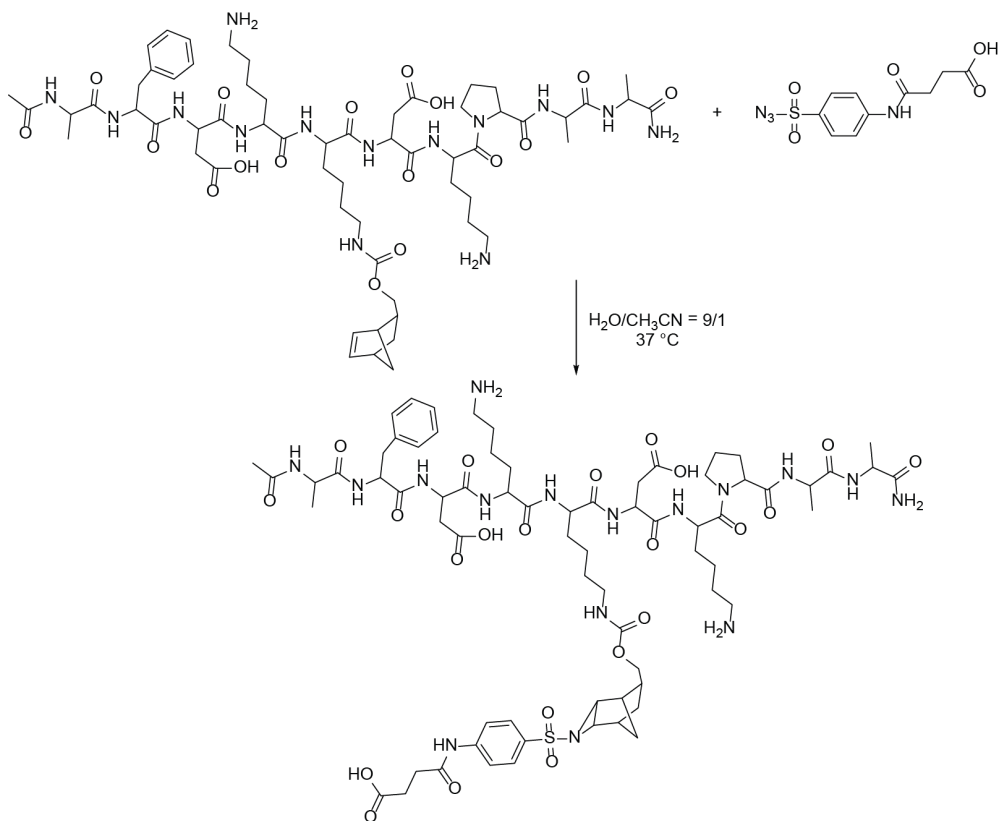
The experiments were performed as follows:

50 μM of the peptide solution was incubated at 37 $^{\circ}\text{C}$ with 2.5 mM 4-((4-(azidosulfonyl)phenyl)amino)-4-oxobutanoic acid **1** in $\text{H}_2\text{O}/\text{CH}_3\text{CN}$ (9:1) mixture and the progress of the reaction was monitored using MALDI mass spectrometry. Probes of the reaction mixture were spotted onto MALDI target using sinapic acid solution as a matrix (SA in 50/50 $\text{H}_2\text{O}/\text{CH}_3\text{CN}$ + 0.1%TFA) and the spectra were recorded in negative mode.

Calculated masses:

Norbornene-containing peptide: $[\text{M}-\text{H}]^-$: 1279.7 Da

Labeled peptide: $[\text{M}-\text{H}]^-$: 1549.7 Da



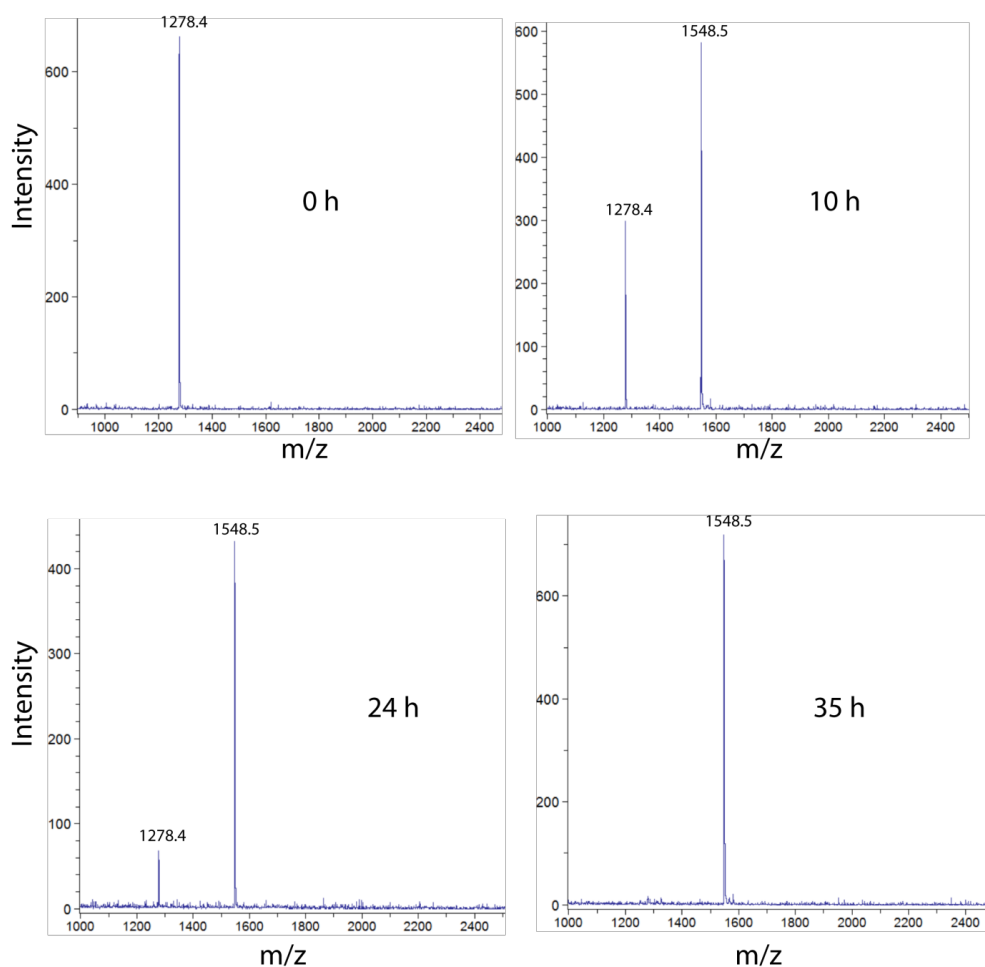


Figure S3: Peptide labeling experiments using 50 μ M peptide and 2.5 mM sulfonyl azide **1**. MALDI-TOF spectra showing the progress of the peptide labeling reaction in time. Starting peptide calc. mass: 1279.7 Da and the product mass: 1549.7 Da; observed: 1278.4 Da and 1548.5 Da respectively. $\Delta M_{\text{calc.}} = 270.0$ Da; $\Delta M_{\text{found}} = 270.1$ Da.

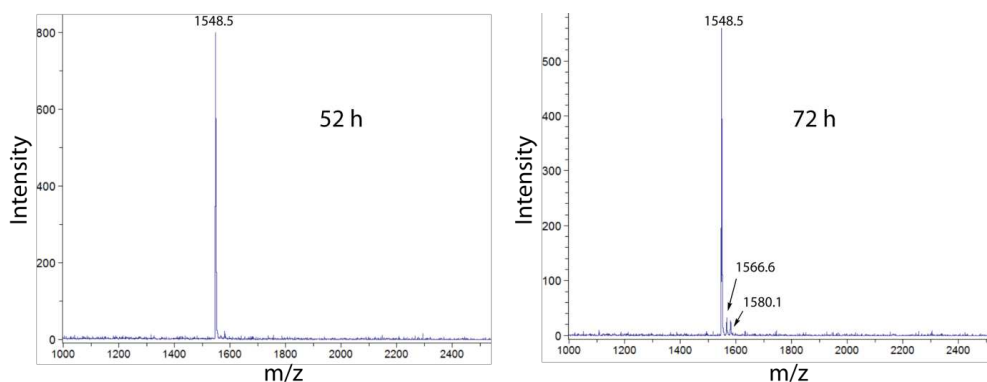


Figure S4: Peptide labeling experiments using 50 μM peptide and 2.5 mM sulfonyl azide **1**. MALDI-TOF spectra showing the formation of the hydrolyzed sulfonamide product after prolonged incubation time (calc. mass $[\text{M}-\text{H}]^-$: 1567.7 Da; observed: 1566.6 Da). The observed mass peak at 1580.1 could not be assigned. Desired aziridine product mass: 1549.7 Da; observed: 1548.5 Da.

Additional experiment using 50 μM peptide and 5 mM (100 equiv.) 4-((4-(azidosulfonyl)phenyl)amino)-4-oxobutanoic acid **1** in $\text{H}_2\text{O}/\text{CH}_3\text{CN}$ (9:1) mixture showed full conversion of the starting peptide to the aziridinated product within 24 hours (Figure S5).

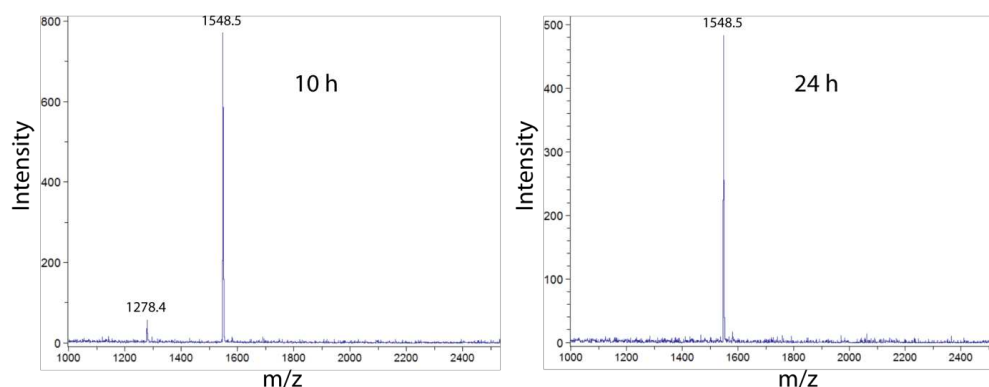


Figure S5: Peptide labeling experiments using 50 μM peptide and 5 mM sulfonyl azide **1**. MALDI-TOF showed full conversion within 24 hours in this case. Starting peptide calc. mass: 1279.7 Da and the product mass: 1549.7 Da; observed: 1278.4 Da and 1548.5 Da respectively. $\Delta M_{\text{calc.}} = 270.0$ Da; $\Delta M_{\text{found}} = 270.1$ Da.

8. Cloning, expression and purification of norbornene-containing proteins

8.1. Cloning and mutagenesis of IBA3_Trx N65amber

The expression vector pPSG_IBA3_trx was created following the Stargate cloning protocol (IBA, Göttingen) using the primers forward *trxA pENTRY_IBA51* and reverse *trxA pENTRY_IBA51* (Table S1). The commercial vector pBAD202 (Life Technologies) was used as template for *trxA* gene amplification. The amber codon (TAG) at position Asn65 was introduced by blunt end site directed mutagenesis using primers *forward trxA N65amber* and *reverse trxA N65amber*.

Table S1: Sequences of the used primers for generation of expression vector IBA trx N65amber.

Name	Sequence
forward <i>trxA pENTRY_IBA51</i>	AGCGGCTCTTC <u>AATG</u> GGA TCT GAT AAA ATT ATT CAT CTG ACT GAT G
reverse <i>trxA pENTRY_IBA51</i>	AGCGGCTCTTC <u>TCCC</u> CAG GTT AGC GTC GAG GAA CTC TTT C
forward <i>trxA N65amber</i>	5'phosph CCG GGC ACT GCG CCG AAA TAT G
reverse <i>trxA N65amber</i>	5'phosph CTA GTG ATC GAT GTT CAG TTT TGC AAC GG

8.2. Expression of norbornene-containing Trx

E. coli strain BL21(DE3) was transformed with pACyc_pylRS Norb, 3xpyIT¹ and pPSG_IBA3_trxA N65amber. An overnight culture was used for inoculation (to OD₆₀₀ of 0.1) of 2 L LB medium containing 2 mM **4**, 50 mg/L carbenicillin and 34 mg/L chloramphenicol. Cells were grown at 37 °C until an OD₆₀₀ of 0.6. Expression of *trxA* N65X was induced by addition 1 mM IPTG. The cells were shaken for further 16 h at 30 °C. After centrifugation (10.000 x g, 10 min, 4°C) the cells were stored at -20°C or directly purified.

All purification steps were carried out at 4°C. Cells were resuspended in StrepA buffer (10 mM Tris-HCl pH 8, 1 mM EDTA, 150 mM NaCl) and supplemented with protease inhibitor mix (*Roche*). The cells were lysed using a French press (*Thermo Scientific*). Cell debris was removed by centrifugation (38 000 x g, 30 min, 4 °C) and the supernatant was applied to a 5 mL Strep-Tactin column (*IBA*), equilibrated with StrepA buffer. Proteins were eluted from the column using the same buffer containing 2.5 mM desthiobiotin. The eluted fractions were pooled and concentrated. Typical protein yields of 25 mg per liter expression medium of **4** containing Trx were achieved. Purified Trx N65X was equilibrated with StrepA buffer, concentrated and stored at -80 °C. A raw intact MS spectrum of Trx N65X (X = norbornene amino acid **4**) is shown in Figure S9. The deconvoluted spectrum is shown in red in Figure 3A of the main text.

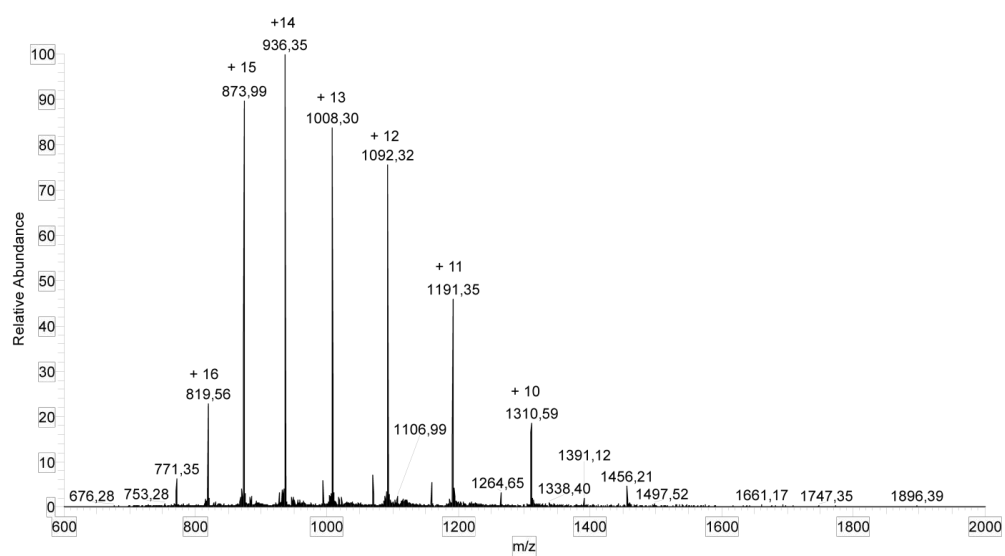


Figure S9: Raw intact MS spectrum of Trx N65X (X = norbornene amino acid 4)

8.3. Tryptic digestion and MS/MS of norbornene-containing Trx

The sequence of Trx is shown in Table S2. Position Asn65 which was chosen for the incorporation of amino acid **4** is shown in red. The peptide generated after tryptic digestion is emphasized in bold letters. Figure S10 shows the corresponding MS/MS spectrum. Table S3 shows the expected and identified MS/MS fragments of the relevant tryptic peptide.

Table S2: Amino acid sequence of Trx

MGSDKIIHLT DDSFDTDVLK ADGAILVDFW AHWCGPCKMI APILDEIADE YQGKLTVAKL
 NIDH**NPGTAP** KYGIRGIPTL LLFKNGEVAA TKVGALSKGQ LKEFLDANLG SAWSHPQFEK

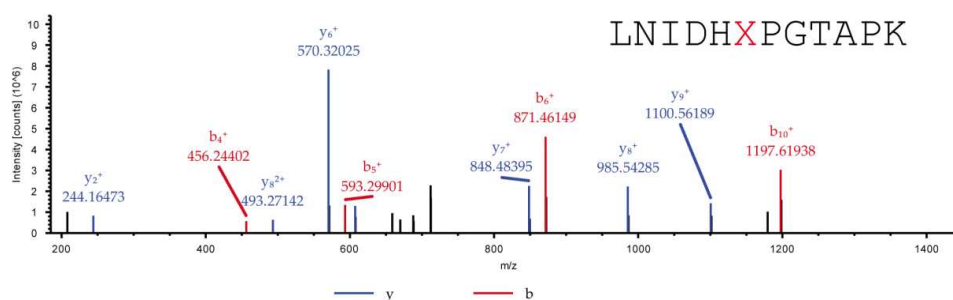


Figure S10: MS/MS spectrum of the tryptic peptide LNIDHXPGTARK (X = **4**). Parent ion: $[M+2H]^{2+}_{\text{calc.}} = 720.8961$, $[M+2H]^{2+}_{\text{obs.}} = 720.8932$ ($\Delta M = 4$ ppm).

Table S3: Expected and identified MS/MS fragments of the tryptic peptide LNIDHXPGTARK (X = **4**). Identified fragments are shown in red for b ions and blue for y ions.

#1	b ⁺	Seq.	y ⁺	y ²⁺	#2
1	114.09135	L			12
2	228.13428	N	1327.70051	664.35389	11
3	341.21835	I	1213.65758	607.33243	10
4	456.24530	D	1100.57351	550.79039	9
5	593.30421	H	985.54656	493.27692	8
6	871.46724	X	848.48765	424.74746	7
7	968.52001	P	570.32462	285.66595	6
8	1025.54148	G	473.27185	237.13956	5
9	1126.58916	T	416.25038	208.62883	4
10	1197.62628	A	315.20270	158.10499	3
11	1294.67905	P	244.16558	122.58643	2
12		K	147.11281	74.06004	1

8.4. Mutagenesis of pACA_HCA H36amber

The amber codon (TAG) was introduced into the expression vector pACA_HCA¹¹ at position His36 of the human carbonic anhydrase II gene by blunt end site directed mutagenesis using the primers *forward HCA H36amber* and *reverse HCA H36amber*.

Table S4: Sequences of the used primers for the generation of expression vector pACA_HCA H36amber.

Name	Sequence
forward HCA H36amber	5'phosph GTT GAC ATC GAC ACT TAG ACA GCC AAG TAT GAC
reverse HCA H36amber	5'phosph AGG GGA CTG GCG CTC TCC CTT GG

8.5. Expression of norbornene containing HCA

The expression vector pACA_HCA G131amber was transformed together with pACyc_pylRS Norb, 3xpyIT¹ which contains the genes of the triple mutant of PylRS and three copies of *pyIT* in *E. coli* BL21(DE3) cells (NEB). 1 L of LB medium containing 34 mg/L chloramphenicol, 100 mg/L carbenicillin and 2 mM norbornene amino acid **4** was inoculated with 10 mL of an overnight culture. The cells were stirred at 37 °C until an OD₆₀₀ of 0.9. At this optical density 1 mM ZnSO₄ and 0.1 mM IPTG were added to induce the expression of the HCA H36amber gene. After further 10 h at 37 °C the cells were harvested and stored at -20 °C until further use. The harvested cells were resuspended in washing buffer (25 mM Tris; 50 mM Na₂SO₄; 50 mM NaClO₄; pH 8.8) and disrupted by French Press procedure. The supernatant of the centrifuged lysate was used for sulfonamide affinity protein purification using an ÄKTA purifier system. The self-packed 3 mL column of *p*-Aminomethylbenzenesulfonamide-Agarose resin (Sigma-Aldrich, A0796) was equilibrated with washing buffer. After binding (0.75 mL/min) of the protein solution, the column was washed with 7 column volumes of washing buffer. HCA was eluted by lowering the pH by elution buffer (100 mM NaOAc; 200 mM NaClO₄; pH 5.6). The protein containing fractions were combined, analyzed by SDS-PAGE, dialyzed against water and lyophilized. Typical yields of the pure norbornene amino acid **4** containing protein HCA H36X were 20 mg/L expression medium.

8.6. Tryptic digestion and MS/MS of norbornene-containing HCA

The sequence of HCA II is shown in Table S5. Position His36 which was chosen for the incorporation of amino acid **4** is shown in red. The peptide generated after tryptic digestion is emphasized in bold letters. Figure S11 shows the corresponding MS/MS spectrum. Table S6 shows the expected and identified MS/MS fragments of the relevant tryptic peptide.

Table S5: Amino acid sequence of HCA II.

10	20	30	40	50	60
MAHHWGYGKH	NGPEHWHKDF	PIAKGER QSP	VDIDT HTAK Y	DPSLKPLSVS	YDQATSLRIL
70	80	90	100	110	120
NNGHAFNVEF	DDSQDKAVLK	GGPLDGTYRL	IQFHFHWGSL	DGQGSEHTVD	KKKYAAELHL
130	140	150	160	170	180
VHWNTKYGDF	CKAVQQPDGL	AVLGIFLKVG	SAKPGLQKVV	DVLDSIKTKG	KSADFTNFDP
190	200	210	220	230	240
RGLLPESLDY	WTPGSLTTP	PLESVTWIV	LKEPISVSSE	QVLKFRKLN	NGEGEPEELM
250	260				
VDNWRPAQPL	KNRQIKASFK				

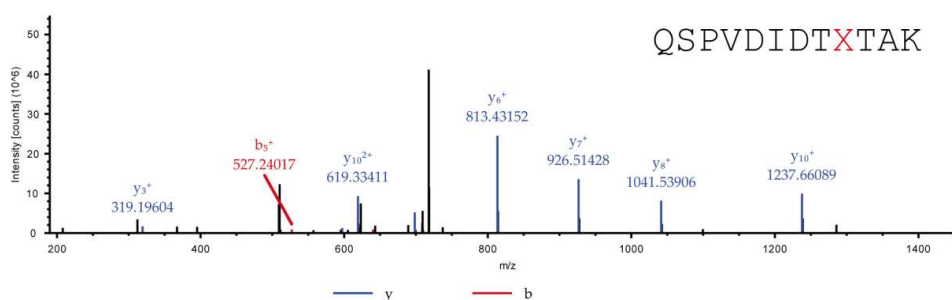


Figure S11: MS/MS spectrum of the tryptic peptide QSPVDIDTHTAK (X = **4**). Parent ion: $[M+2H]^{2+}_{\text{calc.}} = 726.8829$, $[M+2H]^{2+}_{\text{obs.}} = 726.8807$ ($\Delta M = 3$ ppm).

Table S6: Expected and identified MS/MS fragments of the tryptic peptide QSPVDIDTHTAK (X = **4**). Identified fragments are shown in red for b ions and blue for y ions.

#1	b ⁺	Seq.	y ⁺	#2
1	129.06586	Q		12
2	216.09789	S	1324.69949	11
3	313.15066	P	1237.66746	10
4	412.21908	V	1140.61469	9
5	527.24603	D	1041.54627	8
6	640.33010	I	926.51932	7
7	755.35705	D	813.43525	6
8	856.40473	T	698.40830	5
9	1134.56774	X	597.36062	4
10	1235.61542	T	319.19761	3
11	1306.65254	A	218.14993	2
12		K	147.11281	1

9. Protein labeling studies

The purified Trx and HCA proteins were used for modification using different sulfonyl azide reagents. The final concentrations of the proteins were 40 μ M (in 50 mM Tris buffer pH 7.5 for Trx or 50 mM MES buffer pH 5.5 for HCA). The sulfonyl azide derivatives were added to a final concentration of 2 mM. In case of dansyl sulfonyl azide **6** a 10 mM stock solution in DMF was prepared. The final DMF concentration in this case was therefore 20% (v/v). Biotin sulfonyl azide **5** is water soluble and was therefore added without any organic solvent. After overnight incubation at 37 °C the samples were used for SDS-PAGE and/or intact mass spectrometry analysis. To visualize the modification of HCA H36X with the fluorophore **6** the SDS gel was analyzed using an image reader (Fujifilm LAS-3000). The excitation wavelength was 312 nm. The SDS-gels are shown in the main text (Figure 3). The raw intact MS spectrum of Trx labeled with **5** is shown in Figure S12. The deconvoluted spectrum is shown in blue in Figure 3A in the main text.

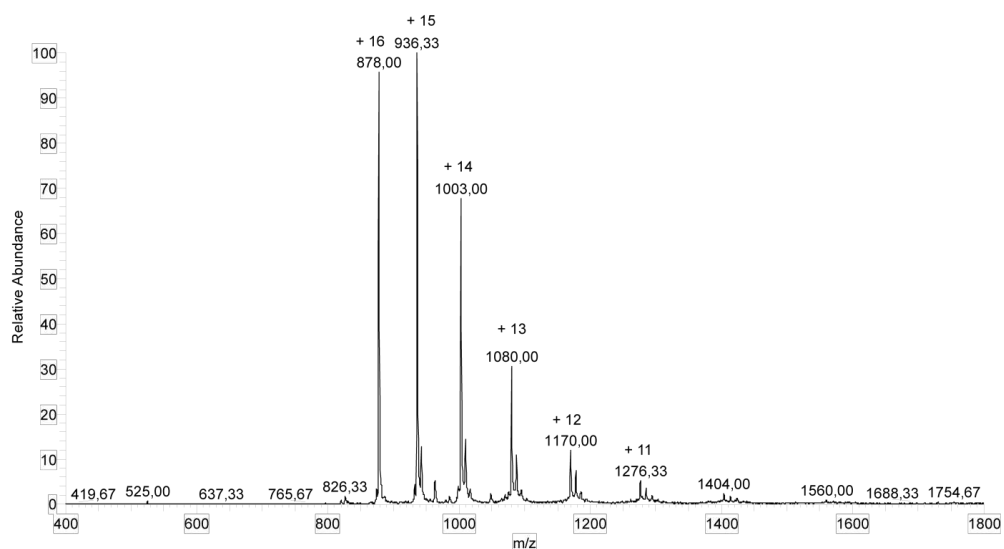


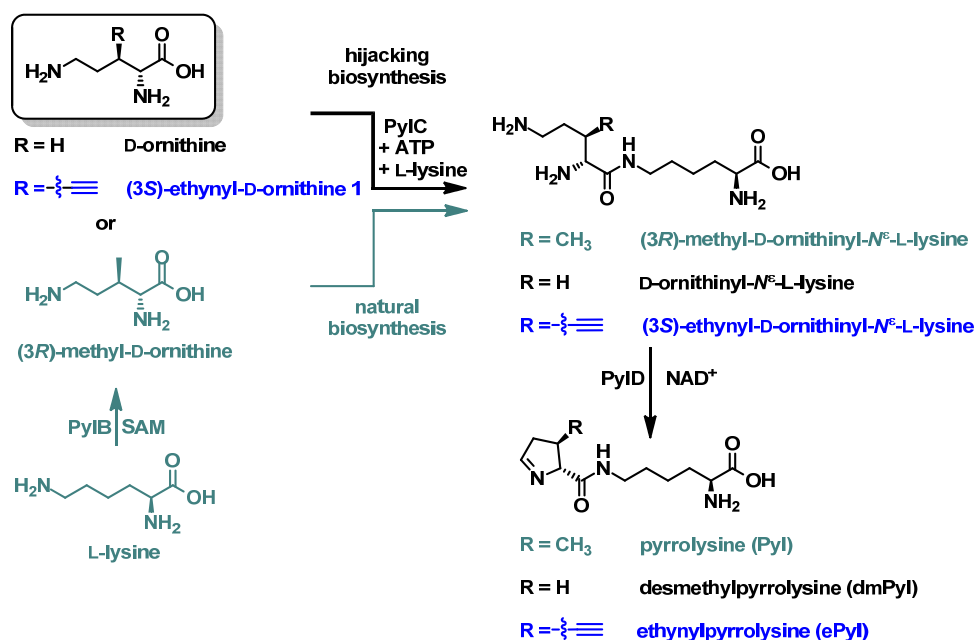
Figure S12: Raw intact MS spectrum of biotinylated Trx N65X (X = norbornene amino acid **4** aziridinated with **5**).

9.3 Selected supplementary information corresponding to section 7.5

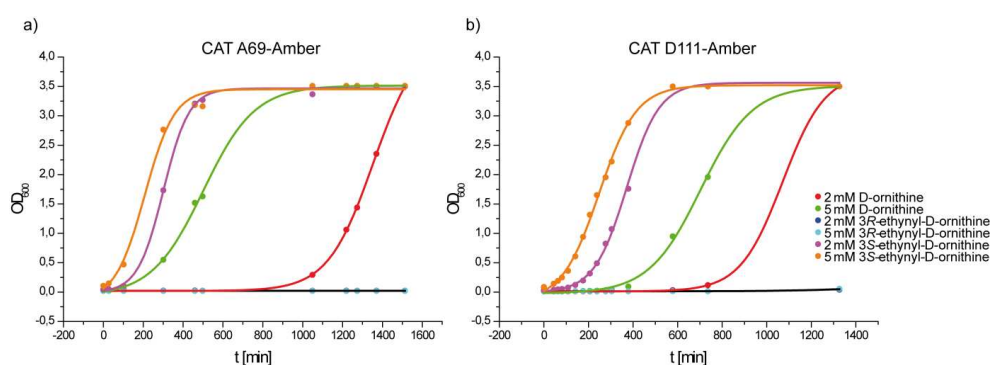
Orchestrating the biosynthesis of an unnatural pyrrolysine amino Acid for its direct incorporation into proteins inside living cells

Pages S-3-S-9 and SI7-S-22 of the supplementary information of the corresponding original publication are shown since they contain methods and data generated by the author of this thesis.

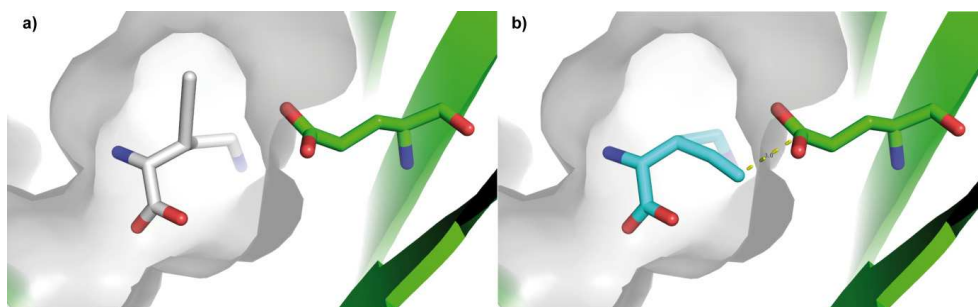
1. Supporting Figures



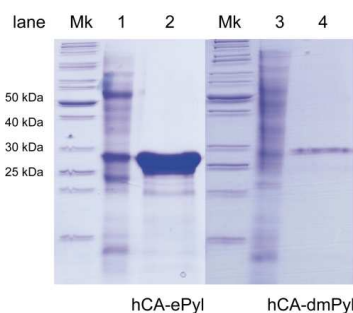
Supporting Figure 1: Scheme of natural and hijacking biosynthetic pathways of pyrrolysine, ethynylpyrrolysine and desmethylpyrrolysine. L-lysine is converted to pyrrolysine via PylB, PylC and PylD (grey) in the natural biosynthesis pathway. 3S-ethynyl-D-ornithine or D-ornithine are converted to ethynylpyrrolysine (blue) and desmethylpyrrolysine (black) respectively via PylC and PylD in the unnatural hitchhiking pathway.



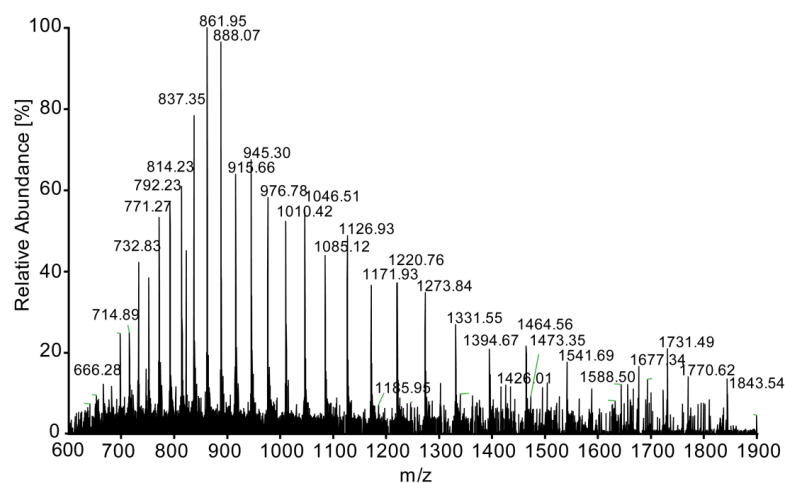
Supporting Figure 2: Biosynthetic conversion of 3S-ethynyl-D-ornithine to ethynylpyrrolysine and incorporation into chloramphenicol acetyltransferase (CAT) using *E. coli* expressing the biosynthesis machinery of Pyl: (PylC, PylD, PylRS, PylT) and CAT-A69amber (a) or CAT-D111amber (b). Growth curves of *E. coli* cells in growth media containing chloramphenicol and supplemented with D-ornithine (2 mM / 5 mM) or 3S-ethynyl-D-ornithine (2 mM / 5 mM) or 3R-ethynyl-D-ornithine (2 mM / 5 mM).



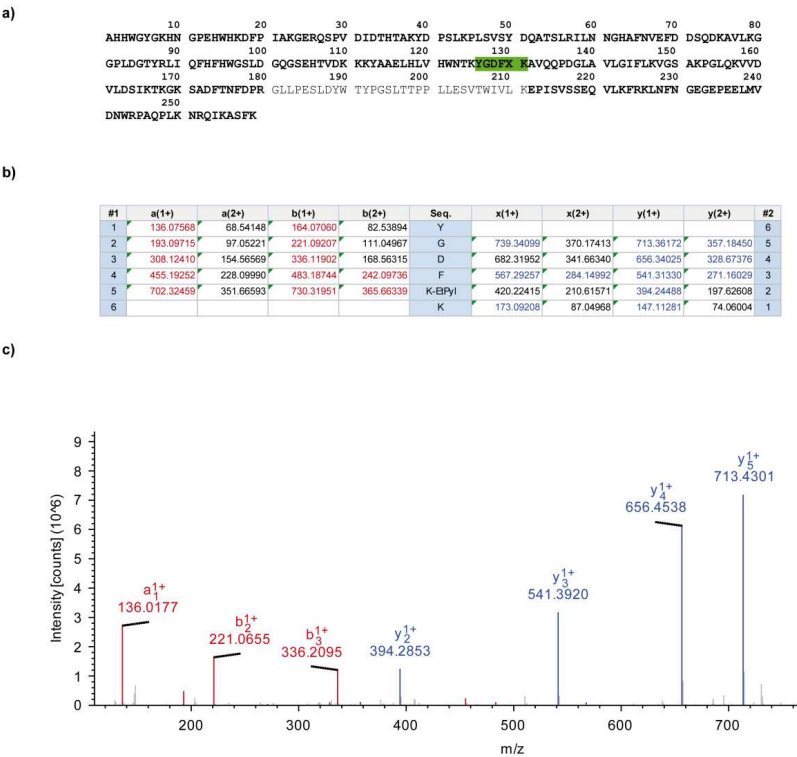
Supporting Figure 3: Molecular Modelling of a) 3*S*-ethynyl-D-ornithine and b) 3*R*-ethynyl-D-ornithine in the catalytic active site of PyIC (PDB:4FFN). In a) the 3*S*-configured ethynyl moiety is located in a hydrophobic pocket whereas in b) 3*R*-configured ethynyl moiety clashes with glutamate residue (Glu227). Molecular modeling was performed with the software *MOLOC*.^[1]



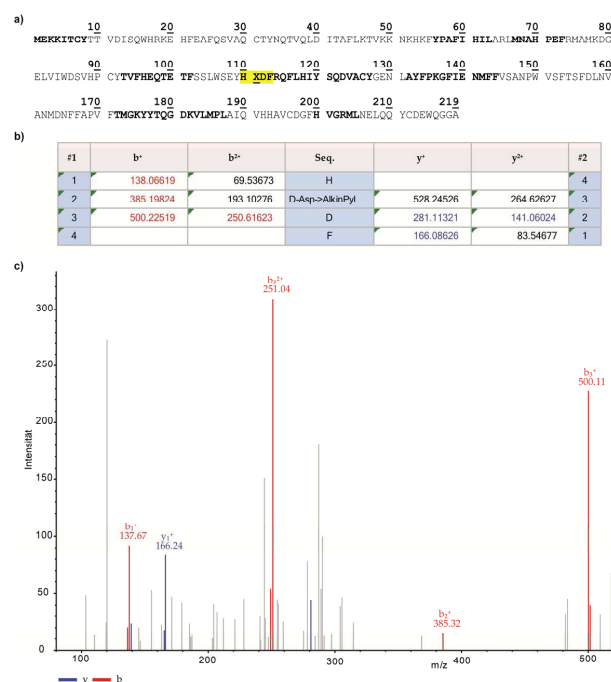
Supporting Figure 4: Expression of hCA-ePyl131 and hCA-dmPyl131. Sample analysis by SDS-PAGE and Coomassie staining. lane 1: lysate of expression of hCA-ePyl131; lane 2: purified hCA-ePyl131; lane 3: lysate of expression of hCA-dmPyl131; lane 4: purified hCA-dmPyl131; Mk: Page Ruler unstained protein ladder (*Thermo*).



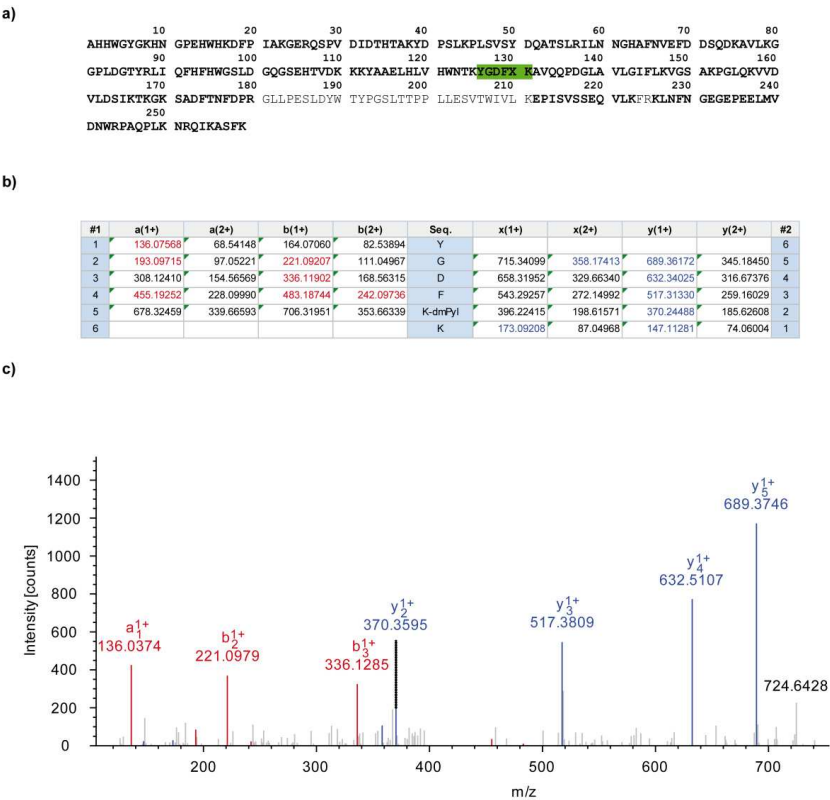
Supporting Figure 5: LC-MS analysis of the incorporated ethynylpyrrolysine in hCA-ePyl131. Intact mass spectrum of hCA-ePyl131 FT-ICR LC-MS measurement.



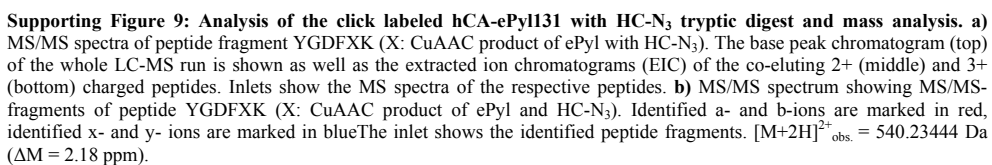
Supporting Figure 6: LC-MS analysis of the incorporated ethynylpyrrolysine in hCA-ePyl131 after tryptic digestion.
a) Amino acid sequence of hCA. Sequence coverage of 88.03% was achieved and identified peptides are depicted in bold letters. Peptide with incorporated ethynylpyrrolysine (X) is marked in green. **b)** Table of calculated and identified fragment ions. Identified b-ions are marked in red, identified y-ions are marked in blue. **c)** MS/MS spectrum showing MS/MS fragments of peptide YGDFXK (X = ePyl). Parent ion: $[M+2H]^{2+}_{\text{obs.}} = 438.71616 \text{ Da}$ ($\Delta M = -0.01 \text{ ppm}$).

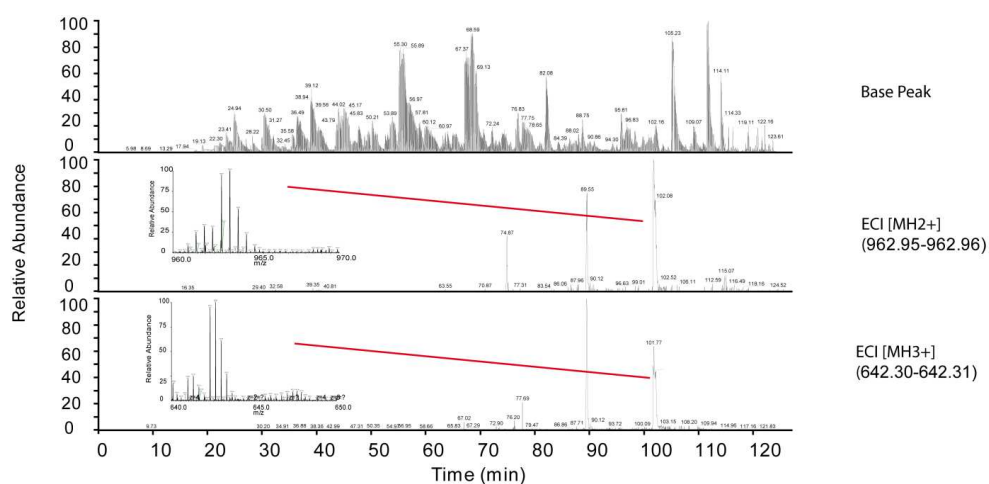


Supporting Figure 7: LC-MS analysis of the incorporated ethynylpyrrolysine in CAT-ePyl111 after chymotryptic digestion. **a)** Amino acid sequence of CAT. 32.88% of the protein were identified. These peptides are depicted in bold letters. The peptide with incorporated ethynylpyrrolysine (X) is marked yellow. **b)** Table of calculated and identified fragment ions. Identified b-ions are marked in red, identified y-ions are marked in blue. **c)** MS/MS spectrum showing MS/MS-fragments of fragment HXDF (X = ePyl). Parent ion: $[M+2H]^{2+}_{\text{obs.}} = 333.15567 \text{ Da}$ ($\Delta M = -0.16 \text{ ppm}$).



Supporting Figure 8: LC-MS analysis of the incorporated desmethylpyrrolysine in hCA-dmPyI131 after tryptic digestion. **a)** Amino acid sequence of hCA. Sequence coverage of 87.26% was achieved and identified peptides are depicted in bold letters. Peptide with incorporated desmethylpyrrolysine (X) is marked in green. **b)** Table of calculated and identified fragment ions. Identified b-ions are marked in red, identified y-ions are marked in blue. **c)** MS/MS spectrum showing MS/MS fragments of peptide YGDFXK (X = dmPyl). Parent ion: $[M+2H]^{2+}_{\text{obs.}} = 426.7166 \text{ Da}$ ($\Delta M = 0.92 \text{ ppm}$).





Supporting Figure 10: Analysis of the double labeled hCA-ePy1131 with HC-N₃ and ABA-bio after tryptic digest and mass analysis. MS/MS spectra of peptide fragment YGDFXX (X: CuAAC and ABA reaction product of ePy1 with HC-N₃ and ABA-bio). The base peak chromatogram (top) of the whole LC-MS run is shown as well as the extracted ion chromatograms (EIC) of the co-eluting 2⁺ (middle) and 3⁺ (bottom) charged peptides. Insets show the MS spectra of the respective peptides. $[M+2H]^{2+}_{\text{obs}} = 962.94875 \text{ Da}$ ($\Delta M = 9.56 \text{ ppm}$).

4. Biochemical procedures

4.1. General remarks

For hCA-ePyl131 expression three plasmids were required. Plasmid 1 contains the Pyl biosynthesis genes *pylC* and *pylD* under the control of an arabinose promoter (*araBAD*) from the pBAD-vector and a Kanamycin resistance gene. Plasmid 2 includes the Pyl-tRNA aminoacyl synthetase gene *pylS* under control of a constitutive *E. coli* glutamin aminoacyl synthetase promoter (*pGlnRS*), three copies of the Pyl tRNA gene (*pylT*) under control of an *E. coli* proK promoter and a chloramphenicol acetyltransferase (*CAT*) gene with an amber codon at position A69 or D111 in a pACyc-vector. Plasmid 3 contains the *hCA* gene with an amber codon at position G131 (*hCA-G131amber*) induced via a T7-promoter on a pACA-vector containing an *AmpR* gene.^[3]

4.2. Incorporation assay

The successful incorporation of ethynylpyrrolysine was verified using the following incorporation assay: The *E. coli* strain BL21(DE3) was transformed with two vectors coding for the pyrrolysine biosynthesis cluster and the chloramphenicol acetyltransferase (*CAT*) with an amber codon (A69O or D111O) in the sequence preliminary to the active site as selection marker (plasmid 1: *pylC* and *pylD*, plasmid 2: *pylS*, 3x *pylT* and *CAT-A69amber* or *CAT-D111amber*).^[4] The LB-medium was supplemented with the antibiotic chloramphenicol (5 ng/μL) and one of the tested amino acids: 3*S*-ethynyl-D-ornithine (5 mM or 2 mM), 3*R*-ethynyl-D-ornithine (5 mM or 2 mM) or D-ornithine (5 mM or 2 mM). The bacteria were grown at 37°C with shaking (200 rpm) and the OD₆₀₀ was recorded at various time points. Cultures supplemented with 3*S*-ethynyl-D-ornithine showed faster growth than cultures supplemented with D-ornithine. Cultures supplemented with 3*R*-ethynyl-D-ornithine showed no increase in the OD₆₀₀ at the recorded time points. Cells from the culture expressing *CAT-D111amber* supplemented with 3*S*-ethynyl-D-ornithine (5 mM) were harvested, sonicated on ice and the lysate was separated using SDS-PAGE. The corresponding *CAT* band was cut out for chymotryptic digestion (18 h, 37 °C) and subjected to mass analysis.

4.3. hCA expression

For the expression of hCA the *E. coli* strain BL21(DE3) was transformed with the described plasmids described in 4.2 and an additional plasmid 3 coding for the hCA-G131amber. LB medium (250 mL) containing arabinose (0.2%) to induce the pyrrolysine biosynthesis cluster, chloramphenicol (15 ng/μL), carbenicillin (50 ng/μL) and either 3*S*-ethynyl-D-ornithine (5 mM) or D-ornithine (5 mM) was inoculated with an overnight culture (2mL) of the transformed bacteria as described previously. The cells were shaken at 37 °C until an OD₆₀₀ of 1.5. At this optical density ZnSO₄ (1 mM) for the active site and IPTG (1 mM) for the expression of the hCA-G131amber gene were added. After 10 h at 37 °C the cells were harvested and stored at -20 °C until further use. The harvested cells were resuspended in wash

buffer (25 mM Tris; 50 mM Na₂SO₄; 50 mM NaClO₄; pH 8.8) supplemented with lysozyme (0.5 mg/L) and lysed by sonication on ice. Then 2.5 U Benzonase (*Novagen*) per mL of original culture was added and centrifuged. The supernatant of the centrifuged lysate was used for sulfonamide affinity protein purification using an ÄKTA purifier system. The self-packed 5 mL column of *p*-aminomethylbenzenesulfonamide-agarose resin (*Sigma-Aldrich*) was equilibrated with hCA wash buffer. After binding (0.75 mL/min) of the lysate, the column was washed with 7 column volumes of hCA wash buffer. hCA was eluted by hCA elution buffer (100 mM NaOAc; 200 mM NaClO₄; pH 5.6). The protein containing fractions were analyzed by SDS-PAGE, combined and rebuffed to phosphate buffered saline (PBS) buffer (137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.76 mM KH₂PO₄; pH 7.4) and used for modification reactions. For long time storage at -80 °C 10% glycerol was added. The correct incorporation of ethynylpyrrolysine or desmethylpyrrolysine was verified *via* LC-MS and LC-MS/MS analysis.

4.4. Protein modification

The versatility of the new incorporated ethynylpyrrolysine was verified in two orthogonal labeling reactions. We conjugated the fluorophore 3-Azido-7-hydroxycoumarin, (*Baseclick*) and/or the synthesized aminobenzaldehyde biotin conjugate ABA-bio with the hCA with ethynylpyrrolysine incorporated at position 131. The coupling reactions were carried out in PBS at pH 7.4 and 25 °C. The *single* conjugation reaction was started by the addition of hCA-ePyl131 (20 µM) to premixed reaction mixture containing the azide (400 µM, stock 4 mM in DMSO) *or* aminobenzaldehyde biotin conjugate (10 mM) and the catalyst master mix CuSO₄ (200 µM), THPTA (500 µM) and freshly dissolved sodium ascorbate (1 mM). For single labeling the reaction mixture was incubated for 16 h. For *double* labeling the aminobenzaldehyde biotin conjugate was added after 2 h incubation time and further incubated for 14 h. The reaction of the aminobenzaldehyde labeling was quenched by adding an aqueous solution of NaCNBH₃ (20 mM) and incubated for 2 h. The samples were rebuffed to (NH₄)₂CO₃ solution (50 mM) using 10 kDa amicon filter ultracel 0.5 mL (*Millipore*). Formation of the protein conjugates were verified by mass spectrometrical analysis (see section mass analysis) and by SDS-PAGE with fluorescence detection using an image reader LAS-3000 (*Fujifilm*) and after transfer to a nitrocellulose membrane and Western Blot using an anti-biotin, HRP-linked antibody (*Cell signaling*) and detection via the SuperSignal West Pico Chemiluminescence Substrate (*Thermo Scientific*).

4.5. Mass analysis

Sample preparation:

For intact protein mass analysis the protein was desalted using 10 kDa Amicon filter ultracel 0.5 mL (*Millipore*) to ddD₂O with 2% MeCN and 0.1% formic acid (FA) and subsequently subjected to mass analysis.

Mass spectrometrical analysis after digestion was either performed directly on the purified protein solution or the visible protein band was cut out of a PAGE-gel. The protein was in-gel alkylated and tryptically or chymotryptically digested.^[5] Trypsin activity was stopped by heatshock and further processed *via* the stage tip method^[6] and analyzed via LC-MS/MS.

LC-MS/MS analysis of peptides:

The samples were analyzed using an UltiMate 3000 nano liquid chromatography system (*Dionex, Fisher Scientific*) coupled to an LTQ-Orbitrap XL (*Fisher Scientific*). For the analysis 12 µl of each sample were injected. The samples were desalted and concentrated on a µ-precolumn cartridge (PepMap100, C18, 5 µM, 100 Å, size 300 µm i.d. x 5 mm) and further processed on a custom made analytical column (ReproSil-Pur, C18, 1.9 µm, 120 Å, packed into a 75 µm i.d. x 150 mm and 8 µm picotip emitter).

The samples were processed via a 127 min multi-step analytical separation at a flow rate 300 nL/min. The gradient with percentages of solvent B was programmed as follows: 1% for 1 minutes; 1%-4% over 1 minute; 4% - 24% over 100 minutes; 24% - 60% over 8 minutes; 60%-85% over 2 minutes; 85% for 5 minutes; 85% - 1% over 2 minutes; 1% for 8 minutes.

Mass spectrometric analyses were performed starting with a full mass scan in the mass range between m/z 150 and 2000 at a resolution of 200000. This survey scan was followed by five scans using the ion trap mass analyzer and CID fragmentation with a normalized collision energy 35. To avoid repeated sequencing of the same peptides, a dynamic exclusion list was used.

Protein identification and proof of incorporation of unnatural amino acids:

The Thermo Proteome Discoverer 1.1 software (*Fisher Scientific*) was used for protein identification. The Sequest search engine was used to identify the proteins from a uniprot “full proteome” data set or the hCA 2 fasta file. As limit of detection a ratio of threefold signal over the noise filter was applied. A maximum of three missed cleavage sites was allowed. The mass tolerances were 20 ppm for the precursor mass and 0.3 Da for the fragment ion mass. As

dynamic modifications ethynylpyrrolysine (K), Pyrrolysine (K) and desmethylpyrrolysine (K) and the corresponding labeling adducts were defined and applied. The integration window tolerance was 15 ppm and the integration method was set to “most confident centroid”.

LC-MS analysis of intact proteins:

The samples were analyzed using an UltiMate 3000 nano liquid chromatography system (Dionex, Fisher Scientific) coupled to a LTQ-FTICR mass spectrometer (Finnigan). 100 ng of each sample were injected for the analysis. The samples were desalted and concentrated on a μ -precolumn cartridge (PepMap300, C4, 5 μ M, 300 Å, size 300 μ m i.d. x 5 mm) and further processed on a custom made analytical column (ReproSil-Pur, C4, 3 μ m, 300 Å, packed into a 75 μ m i.d. x 150 mm and 8 μ m picotip emitter).

A 127 min multi-step analytical separation was performed at a flow rate of 300 nl/min. After holding the gradient on 2% solvent B (acetonitrile containing 0.1% formic acid) and 99% solvent A (water containing 0.1% formic acid) for one minute, it was ramped up to 24% over 100 minutes. After this linear gradient solvent B was increased to 60% over 8 min and subsequent to 85% over 2 min. This level was held for 5 min, then ramped down again to 1% solvent B within 2 min and held for 8 min.

The *Thermo Xcalibur 2.1* and *Protein deconvolution 3.0* software (Fisher Scientific) were used for protein mass analysis and intact mass spectra deconvolution.

5. Literature

- [1] P. R. Gerber, K. Muller, *J. Comput. Aided Mol. Des.* **1995**, *9*, 251-268.
- [2] W. Ou, T. Uno, H. P. Chiu, J. Grunewald, S. E. Cellitti, T. Crossgrove, X. Hao, Q. Fan, L. L. Quinn, P. Patterson, L. Okach, D. H. Jones, S. A. Lesley, A. Brock, B. H. Geierstanger, *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 10437-10442.
- [3] S. Schneider, M. J. Gattner, M. Vrabel, V. Flugel, V. Lopez-Carrillo, S. Prill, T. Carell, *ChemBioChem* **2013**, *14*, 2114-2118.
- [4] a) W. V. Shaw, L. C. Packman, B. D. Burleigh, A. Dell, H. R. Morris, B. S. Hartley, *Nature* **1979**, *282*, 870-872; b) W. V. Shaw, A. G. Leslie, *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *20*, 363-386.
- [5] A. Shevchenko, H. Tomas, J. Havlis, J. V. Olsen, M. Mann, *Nature protocols* **2006**, *1*, 2856-2860.
- [6] N. A. Kulak, G. Pichler, I. Paron, N. Nagaraj, M. Mann, *Nat. Methods* **2014**, *11*, 319-324.